

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 09:14:54 ; Search time 132.195 Seconds
(without alignments)
601.551 Million cell updates/sec

Title: US-09-869-414A-4
Perfect score: 2664
Sequence: 1 MAQALPWLLLWMGAGVLPAA.....CLRCLRQQHDDFADDISLLK 501

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

- 1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
- 23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
- 24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

8						
Result	Query			Description		
No.	Score	Match Length	DB	ID		
1	2664	100.0	501	21	AAU06603	Human aspartyl pro
2	2664	100.0	501	22	AAE10629	Human aspartyl pro
3	2664	100.0	501	22	AAE06859	Human aspartyl pro
4	2664	100.0	501	22	AAU06603	Human Aspartyl pro
5	2664	100.0	501	22	AAU07202	Human aspartyl pro
6	2664	100.0	501	22	AAE02581	Human aspartyl pro
7	2664	100.0	501	23	ABB78590	Human Asp-2(a) pro
8	2664	100.0	501	23	ABB06409	Human aspartyl pro
9	2656	99.7	501	21	AAU94767	Human beta-secreta
10	2656	99.7	501	21	AAB07896	Amino acid sequenc
11	2656	99.7	501	23	ABG78374	Human prepromemaps
12	2656	99.7	509	23	AAM52697	FLAG-tagged human
13	2650	99.5	501	19	AAW59807	Amino acid sequenc
14	2588.5	97.2	969	22	ABG09611	Novel human diagno
15	2582	96.9	488	22	AAB66572	Human memapsin 2.
16	2582	96.9	488	22	AAB61334	Memapsin 2 protein
17	2582	96.9	488	23	ABG78372	Human promemapsin
18	2582	96.9	488	23	AAU99488	Human memapsin 2.
19	2582	96.9	488	24	ABG76101	Human partial mema
20	2582	96.9	501	22	AAB84948	Mouse aspartic sec
21	2582	96.9	503	22	AAB66573	Human pro-memapsin
22	2582	96.9	503	22	AAB61335	T7 promoter and ve
23	2582	96.9	503	23	ABG78373	Human prepromemaps
24	2582	96.9	503	23	AAU99489	Pro-memapsin 2 enc
25	2582	96.9	503	24	ABG76102	Human memapsin 2/T
26	2569	96.4	501	21	AAU94769	Rat beta-secretase
27	2567	96.4	501	21	AAU94768	Murine beta-secret
28	2567	96.4	501	21	AAU94768	Murine aspartyl pr
29	2567	96.4	501	22	AAE10631	Murine aspartyl pr
30	2567	96.4	501	22	AAE06861	Murine aspartyl pr
31	2567	96.4	501	22	AAU06605	Mouse Aspartyl pro
32	2567	96.4	501	22	AAU07204	Mouse aspartyl pro
33	2567	96.4	501	22	AAE02583	Murine aspartyl pr
34	2567	96.4	501	23	ABB78592	Mouse Asp-2(a) pro
35	2506.5	94.1	476	21	AAU94769	Human aspartyl pro
36	2506.5	94.1	476	22	AAE10630	Human aspartyl pro
37	2506.5	94.1	476	22	AAE06860	Human aspartyl pro
38	2506.5	94.1	476	22	AAU06604	Human Aspartyl pro
39	2506.5	94.1	476	22	AAU07203	Human aspartyl pro
40	2506.5	94.1	476	22	AAE02582	Human aspartyl pro
41	2506.5	94.1	476	23	ABB78591	Human Asp-2(b) pro
42	2506.5	94.1	476	23	ABB06410	Human aspartyl pro
43	2506.5	94.1	476	23	ABB06120	Human NS protein s
44	2434	91.4	489	24	ABJ25301	Mouse BACE-interac
45	2420.5	90.9	476	22	AAE06909	Murine aspartyl pr

ALIGNMENTS

RESULT 1

AA88425

ID AA88425 standard; Protein; 501 AA.

XX

AC AA88425;

XX

DT 03-AUG-2000 (first entry)

XX

DE Human aspartyl protease 2 (a) (Asp2) amino acid sequence.

XX

KW Aspartyl protease; aspartase; amyloid precursor protein; APP; Asp 2;

KW Alzheimer's disease; beta secretase site.

XX

OS Homo sapiens.

XX

PN WO200017369-A2.

XX

PD 30-MAR-2000.

XX

PF 23-SEP-1999; 99WO-US20881.

XX

PR 24-SEP-1998; 98US-0101594.

XX

PA (PHAA) PHARMACIA & UPJOHN CO.

XX

PI Gurney ME, Bienkowski MJ, Heinrikson RL, Parodi LA, Yan R;

XX

DR WPI; 2000-303209/26.

DR N-PSDB; AAA15662.

XX

PT New enzyme designated human aspartase useful in research into

PT Alzheimer's Disease is capable of cleaving amyloid protein precursor at

PT the beta secretase site to produce amyloid beta peptide -

XX

PS Claim 48; Fig 2; 183pp; English.

XX

CC This sequence represents the human aspartyl protease 2 (Asp2) amino acid
 CC sequence. The invention relates to a protease (e.g. Asp2) capable of
 CC cleaving the beta secretase site of amyloid precursor protein (APP). The
 CC protease contains a sequence encoding the amino acid sequence DTG and a
 CC sequence encoding DSG or DTG separated by 100-300 amino acids. When
 CC mutated the APP gene causes an autosomal dominant form of Alzheimer's
 CC disease. APP localises to the cell surface membrane and have a single
 CC C-terminal transmembrane domain. Proteolytic processing of APP produces
 CC the amyloid beta protein, which is possibly very important in Alzheimer's
 CC disease. The invention includes a nucleotide sequence encoding the
 CC protease, a vector containing the nucleotide sequence, and a cell line
 CC comprising the vector. Methods for screening for inhibitors of beta
 CC secretase activity are also given in the invention. The human aspartase
 CC protein and nucleotide sequences and the methods for identifying
 CC inhibitors of the protease, are useful in the treatment of and research
 CC in to Alzheimer's disease.

XX

SQ Sequence 501 AA;

Query Match 100.0%; Score 2664; DB 21; Length 501;
 Best Local Similarity 100.0%; Pred. No. 3.8e-263;

	Matches	501;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
Qy	1	MAQALPWLL	WMGAGVLP	PAHGTQH	GIRLPLR	SGLGGAP	LGLRLP	RETDEE	PEEPGR	RGSF 60
Db	1	MAQALPWLL	WMGAGVLP	PAHGTQH	GIRLPLR	SGLGGAP	LGLRLP	RETDEE	PEEPGR	RGSF 60
Qy	61	VEMVDNLR	GKSGQGY	VEMTVGS	PQTLN	ILVDTG	SSNFAV	GAAHPF	LHRY	YQRQLSST 120
Db	61	VEMVDNLR	GKSGQGY	VEMTVGS	PQTLN	ILVDTG	SSNFAV	GAAHPF	LHRY	YQRQLSST 120
Qy	121	YRDLRKG	VYVPYT	QGWEGEL	GTDLV	SIPHGPN	VTVRAN	IAAITE	SDKFF	INGSNWEGIL 180
Db	121	YRDLRKG	VYVPYT	QGWEGEL	GTDLV	SIPHGPN	VTVRAN	IAAITE	SDKFF	INGSNWEGIL 180
Qy	181	GLAYAEI	ARPDDS	LEPFFD	SLVKQ	THVPNL	FSLHLC	GAGFPL	NQSEV	LASVGGSMIIGGI 240
Db	181	GLAYAEI	ARPDDS	LEPFFD	SLVKQ	THVPNL	FSLHLC	GAGFPL	NQSEV	LASVGGSMIIGGI 240
Qy	241	DHSLYTG	SLWYTP	IRREWY	YEVII	VRVEIN	GQDLKM	DCKEYN	YDKSIV	DSGTTNLRPKK 300
Db	241	DHSLYTG	SLWYTP	IRREWY	YEVII	VRVEIN	GQDLKM	DCKEYN	YDKSIV	DSGTTNLRPKK 300
Qy	301	VFEAAVK	SIKAAS	STEKFP	DGFWLG	EQVCWQ	AGTTPW	NIFPVI	SLYLMG	EVTNQSFRT 360
Db	301	VFEAAVK	SIKAAS	STEKFP	DGFWLG	EQVCWQ	AGTTPW	NIFPVI	SLYLMG	EVTNQSFRT 360
Qy	361	ILPQQYL	RPVEDV	ATSQDD	CYKFAI	SQSSTG	TVMGAV	IMEGFY	VVFDRA	RKRIGFAVSAC 420
Db	361	ILPQQYL	RPVEDV	ATSQDD	CYKFAI	SQSSTG	TVMGAV	IMEGFY	VVFDRA	RKRIGFAVSAC 420
Qy	421	HVHDEF	RTAAVE	GPFTLD	MEDCGY	NIPTDE	STLMTI	AYVMAA	ICALFM	LPLCLMVCQW 480
Db	421	HVHDEF	RTAAVE	GPFTLD	MEDCGY	NIPTDE	STLMTI	AYVMAA	ICALFM	LPLCLMVCQW 480
Qy	481	RCLRCLR	QQHDDF	ADDISLLK						501
Db	481	RCLRCLR	QQHDDF	ADDISLLK						501

RESULT 2

AAE10629

ID AAE10629 standard; Protein; 501 AA.

XX

AC AAE10629; ,

XX

DT 10-DEC-2001 (first entry)

XX

DE Human aspartyl protease 2(a) [hu-Asp2(a)] protein.

XX

KW Human; aspartyl protease 2(a); Asp2(a); amyloid precursor protein; APP;

KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;

KW amyloid plaque; neuronal loss; proteolytic; nootropic; neuroprotective;

KW chromosome 11q23.3-24.1.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide 1..21
 FT /label= Signal_peptide
 FT Peptide 22..45
 FT /label= Asp_2a_prepropeptide
 FT Peptide 46..57
 FT /label= Asp_2a_propeptide
 FT Protein 58..501
 FT /label= Mature_human_Asp_2a_protein
 FT Region 420..454
 FT /label= Alpha-helical_spacer_region
 FT Domain 455..477
 FT /label= Transmembrane_domain
 FT Domain 478..501
 FT /label= Cytoplasmic_domain
 XX
 PN GB2357767-A.
 XX
 PD 04-JUL-2001.
 XX
 PF 22-SEP-2000; 2000GB-0023315.
 XX
 PR 23-SEP-1999; 99US-0155493.
 PR 23-SEP-1999; 99US-0404133.
 PR 23-SEP-1999; 99WO-US20881.
 PR 13-OCT-1999; 99US-0416901.
 PR 06-DEC-1999; 99US-0169232.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX
 PI Bienkowski MJ, Gurney M;
 XX
 DR WPI; 2001-444208/48.
 DR N-PSDB; AAD17865.
 XX
 PT Polypeptide comprising fragments of human aspartyl protease with
 PT amyloid precursor protein processing activity and alpha-secretase
 PT activity, for identifying modulators useful in treating Alzheimer's
 PT disease -
 XX
 PS Example 2; Fig 2; 187pp; English.
 XX
 CC The patent discloses human aspartyl protease 1 (hu-Asp1) or modified
 CC Asp1 proteins which lack transmembrane domain or amino terminal
 CC domain or cytoplasmic domain and retains alpha-secretase activity
 CC and amyloid protein precursor (APP) processing activity. The proteins
 CC of the invention are useful for assaying hu-Asp1 alpha-secretase
 CC activity, which in turn is useful for identifying modulators of
 CC hu-Asp1 alpha-secretase activity, where modulators that increase
 CC hu-Asp1 alpha-secretase activity are useful for treating Alzheimer's
 CC disease (AD) which causes progressive dementia with consequent
 CC formation of amyloid plaques, neurofibrillary tangles, gliosis and
 CC neuronal loss. Hu-Asp1 protease substrate is useful for assaying
 CC hu-Asp1 proteolytic activity, by contacting hu-Asp1 protein with
 CC the substrate under acidic conditions and determining the level of
 CC hu-Asp1 proteolytic activity. The present sequence is long form of
 CC human Asp2 protein, designated as Asp2(a). Asp2 gene is localised
 CC on chromosome 11q23.3-24.1.

XX

SQ Sequence 501 AA;

Query Match 100.0%; Score 2664; DB 22; Length 501;
Best Local Similarity 100.0%; Pred. No. 3.8e-263;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MAQALPWL LLWMGAGV LPAHGTQH GIRLPLRSG LGGAPLGLRLP RETDEEPEEPGR RGSF 60
          |||
Db      1 MAQALPWL LLWMGAGV LPAHGTQH GIRLPLRSG LGGAPLGLRLP RETDEEPEEPGR RGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSP PQTLN ILVDTGSSNFAVGAAPHPFLH RYYQRQLSST 120
          |||
Db     61 VEMVDNLRGKSGQGYVEMTVGSP PQTLN ILVDTGSSNFAVGAAPHPFLH RYYQRQLSST 120

Qy    121 YRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNW EGIL 180
          |||
Db    121 YRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNW EGIL 180

Qy    181 GLAYAEIARPD DSLEPF FDSL VKQTHV PNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
          |||
Db    181 GLAYAEIARPD DSLEPF FDSL VKQTHV PNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
          |||
Db    241 DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300

Qy    301 VFEEAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFFVISLYLMGEVTNQSF RIT 360
          |||
Db    301 VFEEAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFFVISLYLMGEVTNQSF RIT 360

Qy    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
          |||
Db    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
          |||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501
          |||
Db    481 RCLRCLRQQHDDFADDISLLK 501
```

RESULT 3

AAE06859

ID AAE06859 standard; Protein; 501 AA.

XX

AC AAE06859;

XX

DT 23-OCT-2001 (first entry)

XX

DE Human aspartyl protease 2a (Hu-Asp2a) protein.

XX

KW Human; aspartyl protease 2a; Asp 2a; beta-amyloid precursor protein; APP;

KW beta-secretase; Alzheimer's disease; dementia; amyloid plaque; gliosis;

KW neurofibrillary tangle; neuronal loss; amyloid-beta peptide; nootropic;

KW neuroprotective; antisense therapy; gene therapy;
 KW chromosome 11q23.3-24.1.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= Signal_peptide
 FT Protein 22..501
 FT /note= "Mature human aspartyl protease 2a (Hu-Asp2a)"
 FT Region 420..454
 FT /note= "Alpha helical spacer region"
 FT Domain 455..477
 FT /label= Transmembrane_domain
 FT Domain 478..501
 FT /label= Cytoplasmic_domain
 XX
 PN WO200150829-A2.
 XX
 PD 19-JUL-2001.
 XX
 PF 09-MAY-2001; 2001WO-IB00799.
 XX
 PR 09-MAY-2001; 2001WO-IB00799.
 XX
 PA (BIEN/) BIENKOWSKI M J.
 PA (GURN/) GURNEY M E.
 PA (HEIN/) HEINRIKSON R L.
 PA (PARO/) PARODI L A.
 PA (YANR/) YAN R.
 XX
 PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;
 XX
 DR WPI; 2001-483072/52.
 DR N-PSDB; AAD13021.
 XX
 PT Novel purified polypeptide comprising fragment of mammalian aspartyl
 PT protease 2, lacking Asp2 transmembrane domain and retaining beta
 PT secretase activity of Asp2 useful for identifying inhibitors of Asp2
 PT activity -
 XX
 PS Claim 49; Fig 2; 185pp; English.
 XX
 CC The invention relates to human aspartyl proteases (Hu-Asp), beta-amyloid
 CC precursor protein (APP) isoforms and their corresponding DNA molecules.
 CC Human aspartyl proteases can act as beta-secretase proteases useful for
 CC treating Alzheimer's disease. APP isoforms are useful for identifying
 CC modulators of amyloid-beta peptide production, for use in designing
 CC therapeutics for the treatment and prevention of Alzheimer's disease,
 CC dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis
 CC and neuronal loss. APP isoforms are also used in methods for identifying
 CC inhibitors and modulators of human Asp2 activity. The invention relates
 CC to a method for identifying agents that modulate the activity of human
 CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used
 CC as a means to screen in cellular assays for the inhibitors of beta- and
 CC gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in
 CC polymerase chain reactions (PCR). The probes are useful for detecting

CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern
CC blots. The present sequence is human aspartyl protease 2 (Hu-Asp2), a
CC 'long' form designated as (Hu-Asp2a). Hu-Asp 2 gene is localised on
CC chromosome 11q23.3-24.1.

XX

SQ Sequence 501 AA;

Query Match 100.0%; Score 2664; DB 22; Length 501;
Best Local Similarity 100.0%; Pred. No. 3.8e-263;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSGLGGAFLGLRLPRETDEEPEEPGRGGSF 60
        |||
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSGLGGAFLGLRLPRETDEEPEEPGRGGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
        |||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy    121 YRDLRKGVPYPTQGWEGELGTDLVSIHPGNVTVRANIAAITESDKFFINGSNWEGIL 180
        |||
Db    121 YRDLRKGVPYPTQGWEGELGTDLVSIHPGNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPDSDLPPFDSLQKTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
        |||
Db    181 GLAYAEIARPDSDLPPFDSLQKTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK 300
        |||
Db    241 DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK 300

Qy    301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
        |||
Db    301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360

Qy    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
        |||
Db    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
        |||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501
        |||
Db    481 RCLRCLRQQHDDFADDISLLK 501
```

RESULT 4

AAU06603

ID AAU06603 standard; Protein; 501 AA.

XX

AC AAU06603;

XX

DT 24-OCT-2001 (first entry)

XX

DE Human Aspartyl protease 2(a), Asp2(a).

XX
 KW Human; Aspartyl protease; Asp2(a); beta-secretase; nootropic;
 KW neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;
 KW amyloid-beta; Abeta.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= Signal peptide
 FT Peptide 22..45
 FT /label= Pre_pro_peptide
 FT Peptide 46..57
 FT /label= Pro_peptide
 FT Protein 57..501
 FT /label= Mature_Asp2(a)
 FT Region 420..454
 FT /label= Alpha_helical_spacer_region
 FT Domian 455..477
 FT /label= Transmembrane_domain
 FT Domian 478..501
 FT /label= Cytoplasmic_domain
 XX
 PN WO200149098-A2.
 XX
 PD 12-JUL-2001.
 XX
 PF 09-MAY-2001; 2001WO-IB00798.
 XX
 PR 09-MAY-2001; 2001WO-IB00798.
 XX
 PA (BIEN/) BIENKOWSKI M J.
 PA (GURN/) GURNEY M E.
 PA (HEIN/) HEINRIKSON R L.
 PA (PARO/) PARODI L A.
 PA (YANR/) YAN R.
 XX
 PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;
 XX
 DR WPI; 2001-502549/55.
 DR N-PSDB; AAS11517.
 XX
 PT Novel purified polypeptide comprising fragment of mammalian aspartyl
 PT protease 2, lacking Asp2 transmembrane domain and retaining beta
 PT secretase activity of Asp2 useful for identifying inhibitors of Asp2
 PT activity -
 XX
 PS Claim 49; Fig 2; 185pp; English.
 XX
 CC The invention relates to a purified polypeptide comprising a fragment of
 CC mammalian aspartyl protease (Asp)2 protein which lacks the Asp2
 CC transmembrane domain and the Asp2 protein, and where the polypeptide and
 CC the fragment retain the beta-secretase activity of the mammalian Asp2
 CC protein. The invention also details polynucleotides for the Asp
 CC proteins and vectors expressing them, and a polypeptide (isoform of
 CC amyloid protein precursor (APP)) comprising the amino acid sequence of an
 CC APP or its fragment containing an APP cleavage site recognizable by a

RESULT 5

AAU07202

ID AAU07202 standard; Protein; 501 AA.

XX

AC AAU07202;

XX

DT 24-OCT-2001 (first entry)

XX

DE Human aspartyl protease 2a (Asp-2a).

XX

KW Human; aspartyl protease 1; Asp-1; nootropic; neuroprotective;

KW aspartyl protease 2; Asp2; amyloid protein precursor; APP;

KW beta-secretase; Alzheimer's disease.

XX

OS Homo sapiens.

XX

FH	Key	Location/Qualifiers
----	-----	---------------------

FT	Peptide	1..21
----	---------	-------

FT		/note= "Signal peptide"
----	--	-------------------------

FT	Misc_feature	22..45
----	--------------	--------

FT		/note= "Pre-propeptide"
----	--	-------------------------

FT	Misc_feature	46..57
----	--------------	--------

FT		/note= "Propeptide"
----	--	---------------------

FT	Protein	58..501
----	---------	---------

FT		/note= "Mature Aspartyl protease-2a"
----	--	--------------------------------------

FT	Region	420..454
----	--------	----------

FT		/note= "Alpha helical spacer region"
----	--	--------------------------------------

FT	Domain	455-477
----	--------	---------

FT		/note= "Transmembrane domain"
----	--	-------------------------------

FT	Domain	478..501
----	--------	----------

FT		/note= "Cytoplasmic domain"
----	--	-----------------------------

XX

PN WO200149097-A2.

XX

PD 12-JUL-2001.

XX

PF 09-MAY-2001; 2001WO-IB00797.

XX

PR 09-MAY-2001; 2001WO-IB00797.

XX

PA (BIEN/) BIENKOWSKI M J.

PA (GURN/) GURNEY M E.

PA (HEIN/) HEINRIKSON R L.

PA (PARO/) PARODI L A.

PA (YANR/) YAN R.

XX

PI Bienkowski MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;

XX

DR WPI; 2001-502548/55.

DR N-PSDB; AAS11702.

XX

PT Novel purified polypeptide comprising fragment of mammalian aspartyl
PT protease 2, lacking Asp2 transmembrane domain and retaining beta
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2
PT activity -


```

Db          361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
Qy          421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
            |||
Db          421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
Qy          481 RCLRCLRQQHDDFADDISLLK 501
            |||
Db          481 RCLRCLRQQHDDFADDISLLK 501

```

RESULT 6

AAE02581

ID AAE02581 standard; Protein; 501 AA.

XX

AC AAE02581;

XX

DT 10-AUG-2001 (first entry)

XX

DE Human aspartyl protease 2a (Asp 2a).

XX

KW Human; alpha-secretase; amyloid precursor protein; APP; therapy;

KW Alzheimer's disease; antialzheimer's; aspartyl protease 2a; Asp 2a;

KW beta-secretase; chromosome 11q23.3-24.1.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Peptide 1..21

FT /label= Signal_peptide

FT Peptide 22..45

FT /label= Asp_2a_prepropeptide

FT Peptide 46..57

FT /label= Asp_2a_propeptide

FT Protein 58..501

FT /label= Mature_human_Asp_2a_protein

FT Active-site 93..95

FT /label= Active_site_1

FT Active-site 289..291

FT /label= Active_site_2

FT Region 420..454

FT /label= Alpha_helical_spacer

FT Domain 455..477

FT /label= Transmembrane_domain

FT Domain 478..501

FT /label= Cytoplasmic_domain

FT Region 486..501

FT /note= "Peptide #2"

XX

PN WO200123533-A2.

XX

PD 05-APR-2001.

XX

PF 22-SEP-2000; 2000WO-US26080.

XX

PR 23-SEP-1999; 99US-0155493.

PR 23-SEP-1999; 99WO-US20881.

PR 13-OCT-1999; 99US-0416901.
PR 06-DEC-1999; 99US-0169232.

XX

PA (PHAA) PHARMACIA & UPJOHN CO.

XX

PI Gurney M, Bienkowski MJ;

XX

DR WPI; 2001-290516/30.

DR N-PSDB; AAD06739.

XX

PT Enzymes that cleave the alpha-secretase site of the amyloid precursor protein, useful for the treatment of Alzheimer's disease -

XX

PS Example 2; Fig 2; 189pp; English.

XX

CC The present invention relates to enzymes for cleaving the alpha-secretase site of the amyloid precursor protein (APP) and methods of identifying those enzymes. The methods may be used to identify enzymes that may be used to cleave the alpha-secretase cleavage site of the APP protein. The enzymes may be used to treat or modulate the progress of Alzheimer's disease. The present sequence is human aspartyl protease 2a (Asp 2a). Asp 2a has beta-secretase protease activity. Asp2 gene is located on chromosome 11q23.3-24.1.

XX

SQ Sequence 501 AA;

Query Match 100.0%; Score 2664; DB 22; Length 501;
Best Local Similarity 100.0%; Pred. No. 3.8e-263;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAQALPWLLWLGAGVLPAGHTQHGI	60
Db	1	MAQALPWLLWLGAGVLPAGHTQHGI	60
Qy	61	VEMVDNLRGKSGQGYVEMTVGSPQTLN	120
Db	61	VEMVDNLRGKSGQGYVEMTVGSPQTLN	120
Qy	121	YRDLRKG VYVPYTQGWEGELGTDLVS	180
Db	121	YRDLRKG VYVPYTQGWEGELGTDLVS	180
Qy	181	GLAYAEIARPDDSLEPFFDSL VKQTH	240
Db	181	GLAYAEIARPDDSLEPFFDSL VKQTH	240
Qy	241	DHSLYTGSLWYTPIRREWYVEIIVR	300
Db	241	DHSLYTGSLWYTPIRREWYVEIIVR	300
Qy	301	VFEAAVKS IKAASSTEKFPDGF	360
Db	301	VFEAAVKS IKAASSTEKFPDGF	360
Qy	361	ILPQQYLRPVEDVATSQDDCYKFAIS	420
Db	361	ILPQQYLRPVEDVATSQDDCYKFAIS	420

Qy 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy 481 RCLRCLRQQHDDFADDISLLK 501
 ||||||||||||||||
 Db 481 RCLRCLRQQHDDFADDISLLK 501

RESULT 7

ABB78590

ID ABB78590 standard; Protein; 501 AA.

XX

AC ABB78590;

XX

DT 16-JUL-2002 (first entry)

XX

DE Human Asp-2(a) protein sequence SEQ ID NO:4.

XX

KW Human; Asp-1; Asp-2; aspartyl protease; enzyme; Alzheimer's disease;
 KW proteolytic; chromosome 11q23.3-24.1.

XX

OS Homo sapiens.

XX

PN GB2367060-A.

XX

PD 27-MAR-2002.

XX

PF 29-OCT-2001; 2001GB-0025934.

XX

PR 23-SEP-1999; 99US-155493P.

PR 23-SEP-1999; 99US-0404133.

PR 23-SEP-1999; 99WO-US20881.

PR 13-OCT-1999; 99US-0416901.

PR 06-DEC-1999; 99US-169232P.

PR 22-SEP-2000; 2000GB-0023315.

XX

PA (PHAA) PHARMACIA & UPJOHN CO.

XX

PI Bienkowski MJ, Gurney M;

XX

DR WPI; 2002-396337/43.

DR N-PSDB; ABL52457.

XX

PT Human aspartyl protease 1 substrates useful in assays to detect
 PT aspartyl protease activity, e.g. for the diagnosis of Alzheimer's
 PT disease -

XX

PS Example 2; Fig 2; 182pp; English.

XX

CC The present invention describes a human aspartyl protease 1 (hu-Asp1)
 CC substrate (I) which comprises a peptide of no more than 50 amino acids,
 CC and which comprises the 8 amino acid sequence Gly-Leu-Ala-Leu-Ala-Leu-
 CC Glu-Pro. Also described are: (1) a method (II) for assaying hu-Asp1
 CC proteolytic activity, comprising: (a) contacting a hu-Asp1 protein with
 CC (I) under acidic conditions; and (b) determining the level of hu-Asp1

CC proteolytic activity; (2) a purified polynucleotide (III) comprising a
 CC nucleotide sequence that hybridises under stringent conditions to the
 CC non-coding strand complementary to a defined 1804 nucleotide sequence
 CC (see ABL52456) where the nucleotide sequence encodes a polypeptide having
 CC Aspl proteolytic activity and lacks nucleotides encoding a transmembrane
 CC domain); (3) a purified polynucleotide (III') comprising a sequence that
 CC hybridises under stringent conditions to (III) (the nucleotide sequence
 CC encodes a polypeptide further lacking a pro-peptide domain corresponding
 CC to amino acids 23-62 of hu-Aspl (see ABB78589)); (4) a vector (IV)
 CC comprising (III) or (III'); and (5) a host cell (V) transformed or
 CC transfected with (III), (III') and/or (IV). The hu-Aspl protease
 CC substrate (I) may be used as an enzyme substrate in assays to detect
 CC aspartyl protease activity, (II) and therefore diagnose diseases
 CC associated with aberrant hu-Aspl expression and activity such as
 CC Alzheimer's disease. Hu-Aspl has been localised to chromosome 21, while
 CC hu-Asp2 has been localised to chromosome 11q23.3-24.1. The present
 CC sequence represents hu-Asp2(a) from the present invention.

XX

SQ Sequence 501 AA;

Query Match 100.0%; Score 2664; DB 23; Length 501;
 Best Local Similarity 100.0%; Pred. No. 3.8e-263;
 Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MAQALPWL	LLWMGAGV	LPAHGTQH	GIRLPLRS	GGLGAPL	GLRLPRET	DEEPEEP	GRGGSF	60
Db	1	MAQALPWL	LLWMGAGV	LPAHGTQH	GIRLPLRS	GGLGAPL	GLRLPRET	DEEPEEP	GRGGSF	60
Qy	61	VEMVDNLR	KGKSGQGY	YVEMTVGS	PQTNLNIL	VDTGSSNF	AVGAAPHF	FLHRYQ	RQLSST	120
Db	61	VEMVDNLR	KGKSGQGY	YVEMTVGS	PQTNLNIL	VDTGSSNF	AVGAAPHF	FLHRYQ	RQLSST	120
Qy	121	YRDLRKGV	YVPYTQGW	EGELGTDL	VSIPHGPN	VTVRANIA	AAITESDK	FFINGSN	WEGIL	180
Db	121	YRDLRKGV	YVPYTQGW	EGELGTDL	VSIPHGPN	VTVRANIA	AAITESDK	FFINGSN	WEGIL	180
Qy	181	GLAYAEIAR	PDDSLEPF	FDLSVKQT	HVPNLFSL	HLCGAGFP	LNQSEVL	ASVGGSM	IIGGI	240
Db	181	GLAYAEIAR	PDDSLEPF	FDLSVKQT	HVPNLFSL	HLCGAGFP	LNQSEVL	ASVGGSM	IIGGI	240
Qy	241	DHSLYTGSL	WYTPIRRE	WYVEVII	VRVEINGQ	DLKMDCKE	YNYDKSIV	DSGTTNLR	LPLPKK	300
Db	241	DHSLYTGSL	WYTPIRRE	WYVEVII	VRVEINGQ	DLKMDCKE	YNYDKSIV	DSGTTNLR	LPLPKK	300
Qy	301	VFEAAVKS	SIKAASST	EKFDPDGF	WLGEQLVC	WQAGTTPW	NIFPVISL	YLMGEVTN	QSFRT	360
Db	301	VFEAAVKS	SIKAASST	EKFDPDGF	WLGEQLVC	WQAGTTPW	NIFPVISL	YLMGEVTN	QSFRT	360
Qy	361	ILPQQYL	LRPVEDV	ATSQDDCY	KFAISQS	STGTVMGA	VIMEGFYV	VFDRARKR	IGFAVS	420
Db	361	ILPQQYL	LRPVEDV	ATSQDDCY	KFAISQS	STGTVMGA	VIMEGFYV	VFDRARKR	IGFAVS	420
Qy	421	HVHDEFRT	AAVEGPFV	TLDMEDCG	YNIPQTDE	STLMTIAY	VMAAICAL	FMLPLCL	MVCQW	480
Db	421	HVHDEFRT	AAVEGPFV	TLDMEDCG	YNIPQTDE	STLMTIAY	VMAAICAL	FMLPLCL	MVCQW	480
Qy	481	RCLRCLRQ	QHDDFADD	ISLLK	501					

Db |||||
481 RCLRCLRQQHDDFADDISLLK 501

RESULT 8

ABB06409

ID ABB06409 standard; Protein; 501 AA.

XX

AC ABB06409;

XX

DT 31-MAY-2002 (first entry)

XX

DE Human aspartyl protease protein sequence SEQ ID NO:2.

XX

KW Beta-secretase; enzyme; cleavage site; amyloid protein precursor; APP;

KW aspartyl protease; neuroprotective; nootropic; beta-secretase inhibitor;

KW Alzheimer's disease.

XX

OS Homo sapiens.

XX

PN WO200206306-A2.

XX

PD 24-JAN-2002.

XX

PF 19-JUL-2001; 2001WO-US23035.

XX

PR 19-JUL-2000; 2000US-219795P.

PR 12-MAR-2001; 2001US-275251P.

XX

PA (PHAA) PHARMACIA & UPJOHN CO.

XX

PI Yan R, Tomasselli AG, Gurney ME, Emmons TL, Bienkowski MJ;

PI Heinrikson RL;

XX

DR WPI; 2002-216995/27.

DR N-PSDB; ABL49914.

XX

PT Novel substrates for human aspartyl protease useful for identifying

PT modulators of beta secretase activity of aspartyl protease for treating

PT Alzheimer's disease -

XX

PS Claim 63; Page 118-119; 188pp; English.

XX

CC The present invention describes an isolated peptide (I) comprising a
CC sequence of at least four amino acids, where the peptide is a substrate
CC for conducting aspartyl protease assays. (I) has neuroprotective and
CC nootropic activities, and can be used as an inhibitor of beta-secretase
CC activity. A beta-secretase modulator from the present invention can be
CC used for inhibiting beta-secretase activity in vivo, and in the
CC manufacture of a medicament for the treatment of Alzheimer's disease.
CC Pharmaceutical compositions from the present invention can be used for
CC treating a disease or condition characterised by an abnormal beta-
CC secretase activity. (I) is useful for identifying agents that modulate
CC the activity of human Asp2 aspartyl protease (Hu-Asp2). (I) is useful
CC as a core structure to construct derivatives. ABL49914 to ABL49925 and
CC ABB06409 to ABB06593 represent sequences used in the exemplification
CC of the present invention.

OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..45
 FT /label= putative signal peptide
 FT Protein 46..501
 FT /label= Beta-secretase
 XX
 PN WO200058479-A1.
 XX
 PD 05-OCT-2000.
 XX
 PF 23-MAR-2000; 2000WO-US07755.
 XX
 PR 26-MAR-1999; 99US-0277229.
 XX
 PA (AMGE-) AMGEN INC.
 XX
 PI Citron M, Vassar RJ, Bennett BD;
 XX
 DR WPI; 2000-594643/56.
 DR N-PSDB; AAA28278.
 XX
 PT Isolated beta-secretase nucleic acids and encoded polypeptides, useful
 PT for diagnosis and gene therapy of Alzheimer's disease -
 XX
 PS Claim 1; Fig 4; 145pp; English.
 XX
 CC This invention relates to 3 nucleotide sequences encoding beta-secretase
 CC proteins. Beta-secretase is an enzyme involved in the production of one
 CC of the components of amyloid plaques involved in Alzheimer's disease. The
 CC invention includes an expression vector comprising the nucleotide
 CC sequence, a host cell comprising the expression vector, and a process for
 CC producing the protein through culturing the transformed cells. Also
 CC included in the invention are a polypeptide derivative of the
 CC beta-secretase protein, a fusion protein comprising beta-secretase fused
 CC to a heterologous amino acid sequence, and a method for modulating the
 CC levels of beta-secretase polypeptide in a mammal comprising administering
 CC the polynucleotide sequence. Beta-secretase exhibits neuroprotective and
 CC nootropic activity. The beta-secretase nucleotide sequence may be used to
 CC map locations of the beta-secretase gene and related genes on chromosomes
 CC and as hybridization probes in diagnostic assays to test for the presence
 CC of beta-secretase DNA or RNA, such as in Alzheimer's disease, Down's
 CC syndrome, and amyloid angiopathy. The nucleotide sequence may also be
 CC used as anti-sense inhibitors of beta-secretase expression, in gene
 CC therapy of Alzheimer's disease, and for the identification of compounds
 CC that modulate beta-secretase activity. Antibodies to the beta-secretase
 CC protein may be used for in vitro and in vivo diagnostic purposes to
 CC detect the presence of beta-secretase polypeptide in a body fluid or cell
 CC sample. The present sequence represents the human beta-secretase protein.
 XX
 SQ Sequence 501 AA;

 Query Match 99.7%; Score 2656; DB 21; Length 501;
 Best Local Similarity 99.8%; Pred. No. 2.5e-262;
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGGSF 60
 |||
 Db 1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGGSF 60

Qy 61 VEMVDNLRGKSGQGYVEMTVGSPPQTLNIIIVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
 |||
 Db 61 VEMVDNLRGKSGQGYVEMTVGSPPQTLNIIIVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy 121 YRDLRKGVPYPTQGKWEDELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180
 |||
 Db 121 YRDLRKGVPYPTQGKWEDELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy 181 GLAYAEIARPDSDLFFDSLVKQTHVPLNLSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
 |||
 Db 181 GLAYAEIARPDSDLFFDSLVKQTHVPLNLSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy 241 DHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPK 300
 |||
 Db 241 DHSLYTGSLWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPK 300

Qy 301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFFVISLYLMGEVTNQSFRT 360
 |||
 Db 301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFFVISLYLMGEVTNQSFRT 360

Qy 361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
 |||
 Db 361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 |||
 Db 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy 481 RCLRCLRQQHDDFADDISLLK 501
 |||
 Db 481 RCLRCLRQQHDDFADDISLLK 501

RESULT 10

AAB07896

ID AAB07896 standard; Protein; 501 AA.

XX

AC AAB07896;

XX

DT 14-NOV-2000 (first entry)

XX

DE Amino acid sequence of a human beta-secretase enzyme.

XX

KW Beta-secretase; beta-amyloid precursor protein; beta-amyloid peptide;
 KW amyloid plaque component; Alzheimer's disease; amyloidogenic disease;
 KW inhibitor.

XX

OS Homo sapiens.

XX

PN WO200047618-A2.

XX

PD 17-AUG-2000.

XX

PF 10-FEB-2000; 2000WO-US03819.

XX

PR 10-FEB-1999; 99US-0119571.

PR 15-JUN-1999; 99US-0139172.

XX

PA (ELAN-) ELAN PHARM INC.

XX

PI Anderson JP, Basi G, Doane MT, Frigon N, John V, Power M;

PI Sinha S, Tatsuno G, Tung J, Wang S, McConlogue L;

XX

DR WPI; 2000-533011/48.

DR N-PSDB; AAA59550, AAA59551.

XX

PT Purified beta-secretase protein used in assays to discover inhibitors

PT which can be used for the treatment of amyloidogenic diseases e.g.

PT Alzheimer's disease -

XX

PS Claim 17; Fig 2A; 121pp; English.

XX

CC The specification describes a beta-secretase enzyme. The enzyme cleaves

CC beta-amyloid precursor protein to produce beta-amyloid peptide. This

CC enzyme is therefore implicated in the production of amyloid plaque

CC components which accumulate in the brains of individuals afflicted with

CC Alzheimer's disease. Inhibitors of beta-secretase are administered to

CC a mammalian subject e.g. with Alzheimer's disease or Alzheimer's

CC disease-like pathology to test if they maintain or improve cognitive

CC ability or reduce the plaque burden. The compounds are used for the

CC treatment of amyloidogenic diseases e.g. Alzheimer's disease. The

CC present sequence represents a human beta-secretase enzyme.

XX

SQ Sequence 501 AA;

Query Match 99.7%; Score 2656; DB 21; Length 501;

Best Local Similarity 99.8%; Pred. No. 2.5e-262;

Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRGGSF 60
          |||
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRGGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQT LNILVD TGSSNFAVGAAPHPFLHRYYQRQLSST 120
          |||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQT LNILVD TGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy    121 YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
          |||
Db    121 YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPD DSLEPFFDSL VKQTHVPNL FSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
          |||
Db    181 GLAYAEIARPD DSLEPFFDSL VKQTHVPNL FSLQLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHS LYTGS LWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
          |||
Db    241 DHS LYTGS LWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300

Qy    301 VFEAAVKS IKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360
```

```

      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 VFEEAAVKSIIKAASSTEFKPDGFWLGEQLVCWQAGTTPWNIFFVISLYLMGEVTNQSFRT 360
Qy      361 ILPQQYLRLPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 ILPQQYLRLPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
Qy      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
Qy      481 RCLRCLRQQHDDFADDISLLK 501
      ||||||||||||||||||
Db      481 RCLRCLRQQHDDFADDISLLK 501

```

RESULT 11

ABG78374

ID ABG78374 standard; Protein; 501 AA.

XX

AC ABG78374;

XX

DT 15-NOV-2002 (first entry)

XX

DE Human prepromemapsin 2.

XX

KW Human; enzyme; memapsin 2; aspartic protease; beta secretase;
 KW degenerative disease; Alzheimer's disease; amyloid precursor protein;
 KW APP; neuroprotective; nootropic; inhibitor;
 KW substrate side-chain preference.

XX

OS Homo sapiens.

XX

PN WO200253594-A2.

XX

PD 11-JUL-2002.

XX

PF 28-DEC-2001; 2001WO-US50826.

XX

PR 28-DEC-2000; 2000US-258705P.

PR

14-MAR-2001; 2001US-275756P.

XX

PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.

PA

(UNII) UNIV ILLINOIS FOUND.

XX

PI Tang JJN, Koelsch G, Ghosh AK;

XX

DR WPI; 2002-619088/66.

XX

PT New memapsin 2 activity inhibitor useful in treatment of e.g.

PT

Alzheimer's disease -

XX

PS Disclosure; Fig 9; 74pp; English.

XX

CC The invention relates to an inhibitor of catalytically active memapsin 2
 CC (an aspartic protease which can cleave at beta secretase sites), which
 CC binds to the active site of memapsin 2 defined by the presence of two

CC catalytic aspartic residues and substrate binding cleft. Also
 CC included is a method of determination of the substrate side-chain
 CC preference in memapsin 2 sub-sites comprising: (a) reacting a mixture of
 CC memapsin 2 substrates with memapsin 2, and determining the sub-site
 CC preference of memapsin 2 by determining relative initial hydrolysis rates
 CC of the mixture of memapsin 2 substrates; or (b) preparing a combinatorial
 CC library of memapsin 2 inhibitors containing a base sequence taken from
 CC OM99-2 (Glu-Val-Asn-Leu-Ala-Ala-Glu-phe), probing the library of
 CC inhibitors with memapsin 2 which binds to several inhibitors to generate
 CC several bound memapsin 2, and detecting the bound memapsin 2 with an
 CC antibody raised to memapsin 2 and an alkaline phosphatase conjugated
 CC secondary antibody. The inhibitors may be used in the manufacture of a
 CC medicament for the treatment of Alzheimer's disease since memapsin 2 may
 CC be involved in the cleavage of amyloid precursor protein (APP), and for
 CC determining the substrate side-chain preference in memapsin 2 sub-sites.
 CC The present sequence represents human memapsin 2 (either prepromemapsin 2
 CC or mature memapsin).

XX

SQ Sequence 501 AA;

Query Match 99.7%; Score 2656; DB 23; Length 501;
 Best Local Similarity 99.8%; Pred. No. 2.5e-262;
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGGSF 60
      |||
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQTINILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
      |||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQTINILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy    121 YRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
      |||
Db    121 YRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPDSDLPPFDSLKVQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
      |||
Db    181 GLAYAEIARPDSDLPPFDSLKVQTHVPNLFSLQLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK 300
      |||
Db    241 DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK 300

Qy    301 VFEEAAVKSIIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
      |||
Db    301 VFEEAAVKSIIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360

Qy    361 ILPQQYLRLPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
      |||
Db    361 ILPQQYLRLPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALEFMLPLCLMVCQW 480
      |||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALEFMLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501

```

Db |||||
481 RCLRCLRQQHDDFADDISLLK 501

RESULT 12

AAM52697

ID AAM52697 standard; Protein; 509 AA.

XX

AC AAM52697;

XX

DT 26-FEB-2002 (first entry)

XX

DE FLAG-tagged human beta-secretase.

XX

KW Human; beta-secretase; FLAG tag; inhibitor; amine compound;
KW beta amyloid protein production; head injury; spinal injury;
KW amyloid precursor protein alpha secretion; nerve damage;
KW meningitis sequela; cerebral paralysis; memory disorder;
KW mental disease; nootropic; neuroprotective; cerebroprotective.

XX

OS Homo sapiens.

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Region 502..509

FT /label= FLAG_tag

XX

PN WO200187293-A1.

XX

PD 22-NOV-2001.

XX

PF 18-MAY-2001; 2001WO-JP04144.

XX

PR 19-MAY-2000; 2000JP-0152758.

XX

PA (TAKE) TAKEDA CHEM IND LTD.

XX

PI Miyamoto M, Matsui J, Fukumoto H, Tarui N;

XX

DR WPI; 2002-055640/07.

DR N-PSDB; ABA02406.

XX

PT Beta-secretase inhibitor used for treating e.g. Alzheimer's disease and
PT injury to brain or spine, and neurodegeneration, comprises amine
PT compound -

XX

PS Examples; Page 79-81; 86pp; Japanese.

XX

CC The invention relates to novel amine compounds which are beta-secretase
CC inhibitors. The beta-secretase compounds also have the ability to
CC promote amyloid precursor protein alpha secretion and to inhibit beta
CC amyloid protein production. The beta-secretase inhibitors of the
CC invention can be used for treating head or spinal injuries, nerve damage,
CC sequelae of meningitis, cerebral paralysis, memory disorders and mental
CC diseases. The present sequence represents a FLAG-tagged human
CC beta-secretase used in the exemplifications of the invention.

XX

SQ Sequence 509 AA;

Query Match 99.7%; Score 2656; DB 23; Length 509;
Best Local Similarity 99.8%; Pred. No. 2.6e-262;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MAQALPWLLWLGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRGSF 60
        |||||||
Db      1 MAQALPWLLWLGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQTLN I LVD TGSSNFAVGAAPHF L H R Y Y Q R Q L S S T 120
        |||||||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQTLN I LVD TGSSNFAVGAAPHF L H R Y Y Q R Q L S S T 120

Qy    121 YRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWE G I L 180
        |||||||
Db    121 YRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWE G I L 180

Qy    181 GLAYAEIARPD DSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
        |||||||
Db    181 GLAYAEIARPD DSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWY YEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR L P K K 300
        |||||||
Db    241 DHSLYTGSLWYTPIRREWY YEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR L P K K 300

Qy    301 VFEEAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF R I T 360
        |||||||
Db    301 VFEEAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF R I T 360

Qy    361 ILPQQYL R P V E D V A T S Q D D C Y K F A I S Q S S T G T V M G A V I M E G F Y V V F D R A R K R I G F A V S A C 420
        |||||||
Db    361 ILPQQYL R P V E D V A T S Q D D C Y K F A I S Q S S T G T V M G A V I M E G F Y V V F D R A R K R I G F A V S A C 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALEFMLPLCLMVCQW 480
        |||||||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALEFMLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501
        |||||||
Db    481 RCLRCLRQQHDDFADDISLLK 501
```

RESULT 13

AAW59807

ID AAW59807 standard; Protein; 501 AA.

XX

AC AAW59807;

XX

DT 26-OCT-1998 (first entry)

XX

DE Amino acid sequence of human ASP2 (aspartic protease 2).

XX

KW Human; ASP2; aspartic protease 2; agonist; antagonist; immunospecific;
KW antibody; inhibition; Alzheimer's disease; cancer; proteinase;
KW prohormone processing.

XX

```

OS   Homo sapiens.
XX
PN   EP855444-A2.
XX
PD   29-JUL-1998.
XX
PF   27-JAN-1998;   98EP-0300573.
XX
PR   28-JAN-1997;   97GB-0001684.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
PA   (SMIK ) SMITHKLINE BEECHAM PLC.
XX
PI   Chapman CG,  Murphy K,  Powell DJ,  Smith TS;
XX
DR   WPI; 1998-389809/34.
DR   N-PSDB; AAV41696.
XX
PT   New nucleic acid encoding human aspartic protease 2 - used to treat,
PT   prevent and diagnose e.g. Alzheimer's disease, cancer and prohormone
PT   processing
XX
PS   Claim 1; Page 7; 26pp; English.
XX
CC   This is the amino acid sequence of the human ASP2 (aspartic protease
CC   family), used in the method of the invention. Agonists and
CC   antagonists for ASP2 immunospecific antibodies are used to treat
CC   conditions requiring increased or decreased activity or expression of
CC   ASP2 respectively. ASP2 is used to treat and diagnose e.g.
CC   Alzheimer's disease, cancer and prohormone processing and ASP2 or a
CC   fragment can be used to induce an immune response against the above
CC   conditions.
XX
SQ   Sequence    501 AA;

Query Match          99.5%; Score 2650; DB 19; Length 501;
Best Local Similarity 99.6%; Pred. No. 1e-261;
Matches 499; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGSF 60
        |||
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGSF 60

Qy      61 VEMVDNLRGKSGQGYVEMTVGSPQTLNIIIVDTGSSNFAVGAAPHPFLHRYYQRLSST 120
        |||
Db      61 VEMVDNLRGKSGQGYVEMTVGSPQTLNIIIVDTGSSNFAVGAAPHPFLHRYYQRLSST 120

Qy      121 YRDLRKGVPYPYTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180
        |||
Db      121 YRDLRKGVPYPYTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy      181 GLAYAEIARPDDSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
        |||
Db      181 GLAYAEIARPDDSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy      241 DHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRLPKK 300
        |||

```


CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX
 SQ Sequence 969 AA;

Query Match 97.2%; Score 2588.5; DB 22; Length 969;
 Best Local Similarity 98.0%; Pred. No. 6e-255;
 Matches 492; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy	1	MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEEPEEPGRGGSF	60
Db	1	MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEEPEEPGRGGSF	60
Qy	61	VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAHPFLHRYRQQLSST	120
Db	61	VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAHPFLHRYRQQLSST	120
Qy	121	YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL	180
Db	121	YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL	180
Qy	181	GLAYAEIARPD DSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI	240
Db	181	GLAYAEIARPD DSLEPFFDSL VKQTHVPNLFSLQLCGAGFPLNQSEVLASVGGSMIIGGI	240
Qy	241	DHSLYTGSLWYTPIRREWYVEV IIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK	300
Db	241	DHSLYTGSLWYTPIRRESYVEV IIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK	300
Qy	301	VFEAAVKS IKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT	360
Db	301	VFEAAVKS IKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT	360
Qy	361	ILPQQYLRPVEDVAT SQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC	420
Db	361	ILPQQYLRPVEDVAT SQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGLLVSAC	420
Qy	421	HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW	480

```

Db      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
Qy      481 RCLRCLRQQHD-DFADDISLLK 501
        |||||
Db      481 RCLRCLRQQHGMTLPDDISLLK 502

```

RESULT 15

AAB66572

ID AAB66572 standard; Protein; 488 AA.

XX

AC AAB66572;

XX

DT 12-APR-2001 (first entry)

XX

DE Human memapsin 2.

XX

KW Human; memapsin 2; nootropic; neuroprotective; amyloid precursor protein;

KW APP; memapsin 2 inhibitor; Alzheimer's disease.

XX

OS Homo sapiens.

XX

PN WO200100665-A2.

XX

PD 04-JAN-2001.

XX

PF 27-JUN-2000; 2000WO-US17742.

XX

PR 28-JUN-1999; 99US-0141363.

PR 30-NOV-1999; 99US-0168060.

PR 25-JAN-2000; 2000US-0177836.

PR 27-JAN-2000; 2000US-0178368.

PR 08-JUN-2000; 2000US-0210292.

XX

PA (OKLA-) OKLAHOMA MEDICAL RES FOUND.

PA (UNII) UNIV ILLINOIS FOUND.

XX

PI Tang JJN, Hong L, Ghosh AK;

XX

DR WPI; 2001-137933/14.

DR N-PSDB; AAF31848.

XX

PT Novel memapsin 2 inhibitors which bind to active site of memapsin 2

PT having 2 catalytic aspartic residues and substrate binding cleft, used

PT to treat Alzheimer's disease by blocking amyloid precursor protein

PT cleavage -

XX

PS Example 1; Page 72-74; 86pp; English.

XX

CC The present sequence is given in a specification relating to an inhibitor
 CC of catalytically active memapsin 2. The inhibitor binds to the memapsin 2
 CC active site, which is defined by the presence of two catalytic aspartic
 CC residues and a substrate binding cleft. The inhibitor is useful for
 CC the treatment and diagnosis of Alzheimer's disease. It is useful in
 CC screens for individuals with a genetic predisposition to Alzheimer's
 CC disease. The inhibitor is useful as a reagent for specifically binding to

CC memapsin 2 or memapsin 2 analogues and for aiding in memapsin 2
CC isolation, purification and characterisation.
XX
SQ Sequence 488 AA;

Query Match 96.9%; Score 2582; DB 22; Length 488;
Best Local Similarity 99.8%; Pred. No. 9e-255;
Matches 487; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      14 AGVLP AHGTQH GIRLPLR SGLGGAPLGLRLPRETDEEPEEPGRRGSFVEMVDNLRGKSGQ 73
          |||
Db       1 AGVLP AHGTQH GIRLPLR SGLGGAPLGLRLPRETDEEPEEPGRRGSFVEMVDNLRGKSGQ 60

Qy      74 GYYVEMTVGSP PQTLN ILVDTGSSNFAVGAAPHPFLHRYYQRQLSSTYRDLRKG VYVPYT 133
          |||
Db      61 GYYVEMTVGSP PQTLN ILVDTGSSNFAVGAAPHPFLHRYYQRQLSSTYRDLRKG VYVPYT 120

Qy     134 QGKWE GELGTDLV SIPHGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDDS 193
          |||
Db     121 QGKWE GELGTDLV SIPHGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDDS 180

Qy     194 LEPFFDSL VKQTHVP NLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSLWYTP 253
          |||
Db     181 LEPFFDSL VKQTHVP NLFSLQLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSLWYTP 240

Qy     254 IRREWYYEVI IVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKKVFEAAVKS IKAAS 313
          |||
Db     241 IRREWYYEVI IVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKKVFEAAVKS IKAAS 300

Qy     314 STEKF PDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RITILPQQYLRPVEDV 373
          |||
Db     301 STEKF PDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RITILPQQYLRPVEDV 360

Qy     374 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSACHVHDEFRTAAVEG 433
          |||
Db     361 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSACHVHDEFRTAAVEG 420

Qy     434 PFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQWRCLRCLRQQHDDF 493
          |||
Db     421 PFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQWRCLRCLRQQHDDF 480

Qy     494 ADDISLLK 501
          |||
Db     481 ADDISLLK 488
```

Search completed: January 21, 2004, 09:22:24
Job time : 133.195 secs

OM protein - protein search, using sw model

Run on: January 21, 2004, 09:19:55 ; Search time 45.0229 Seconds
(without alignments)
470.821 Million cell updates/sec

Title: US-09-869-414A-4
Perfect score: 2664
Sequence: 1 MAQALPWLLWLGAGVLPAA.....CLRCRLRQQHDDFADDISLLK 501

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	2656	99.7	501	4	US-09-548-372D-4	Sequence 4, Appli
2	2656	99.7	501	4	US-09-548-367D-4	Sequence 4, Appli
3	2656	99.7	501	4	US-09-551-853D-4	Sequence 4, Appli
4	2650	99.5	501	4	US-09-009-191-2	Sequence 2, Appli
5	2582	96.9	488	4	US-09-604-608-2	Sequence 2, Appli
6	2582	96.9	501	4	US-09-713-158-2	Sequence 2, Appli
7	2582	96.9	503	4	US-09-604-608-3	Sequence 3, Appli
8	2567	96.4	501	4	US-09-548-372D-8	Sequence 8, Appli
9	2567	96.4	501	4	US-09-548-367D-8	Sequence 8, Appli
10	2567	96.4	501	4	US-09-551-853D-8	Sequence 8, Appli
11	2506.5	94.1	476	4	US-09-548-372D-6	Sequence 6, Appli

12	2506.5	94.1	476	4	US-09-548-367D-6	Sequence 6, Appli
13	2506.5	94.1	476	4	US-09-551-853D-6	Sequence 6, Appli
14	2420.5	90.9	476	4	US-09-548-372D-73	Sequence 73, Appl
15	2420.5	90.9	476	4	US-09-548-367D-73	Sequence 73, Appl
16	2420.5	90.9	476	4	US-09-551-853D-73	Sequence 73, Appl
17	2397	90.0	453	4	US-09-548-372D-30	Sequence 30, Appl
18	2397	90.0	453	4	US-09-548-367D-30	Sequence 30, Appl
19	2397	90.0	453	4	US-09-551-853D-30	Sequence 30, Appl
20	2397	90.0	459	4	US-09-548-372D-32	Sequence 32, Appl
21	2397	90.0	459	4	US-09-548-367D-32	Sequence 32, Appl
22	2397	90.0	459	4	US-09-551-853D-32	Sequence 32, Appl
23	2315	86.9	774	4	US-09-009-191-4	Sequence 4, Appli
24	2291.5	86.0	446	4	US-09-548-372D-22	Sequence 22, Appl
25	2291.5	86.0	446	4	US-09-548-367D-22	Sequence 22, Appl
26	2291.5	86.0	446	4	US-09-551-853D-22	Sequence 22, Appl
27	2288	85.9	433	4	US-09-548-372D-26	Sequence 26, Appl
28	2288	85.9	433	4	US-09-548-367D-26	Sequence 26, Appl
29	2288	85.9	433	4	US-09-551-853D-26	Sequence 26, Appl
30	2288	85.9	459	4	US-09-548-372D-24	Sequence 24, Appl
31	2288	85.9	459	4	US-09-548-367D-24	Sequence 24, Appl
32	2288	85.9	459	4	US-09-551-853D-24	Sequence 24, Appl
33	2247.5	84.4	428	4	US-09-548-372D-51	Sequence 51, Appl
34	2247.5	84.4	428	4	US-09-548-367D-51	Sequence 51, Appl
35	2247.5	84.4	428	4	US-09-551-853D-51	Sequence 51, Appl
36	2247.5	84.4	434	4	US-09-548-372D-53	Sequence 53, Appl
37	2247.5	84.4	434	4	US-09-548-367D-53	Sequence 53, Appl
38	2247.5	84.4	434	4	US-09-551-853D-53	Sequence 53, Appl
39	2104	79.0	425	4	US-09-548-372D-28	Sequence 28, Appl
40	2104	79.0	425	4	US-09-548-367D-28	Sequence 28, Appl
41	2104	79.0	425	4	US-09-551-853D-28	Sequence 28, Appl
42	1173.5	44.1	518	3	US-08-999-723-2	Sequence 2, Appli
43	1173.5	44.1	518	3	US-09-434-427-2	Sequence 2, Appli
44	1173.5	44.1	518	4	US-09-548-372D-2	Sequence 2, Appli
45	1173.5	44.1	518	4	US-09-548-367D-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-09-548-372D-4

; Sequence 4, Application US/09548372D

; Patent No. 6420534

; GENERAL INFORMATION:

; APPLICANT: GURNEY ET AL.

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: 29915/6280I

; CURRENT APPLICATION NUMBER: US/09/548,372D

; CURRENT FILING DATE: 2000-04-12

; PRIOR APPLICATION NUMBER: US 60/155,493

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 09/404,133

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: PCT/US99/20881

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-548-372D-4

Query Match 99.7%; Score 2656; DB 4; Length 501;
Best Local Similarity 99.8%; Pred. No. 7.4e-267;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MAQALPWL...LWMGAGVLP...AHGTQHGIRLPLR...SGLGGAPLGLRLPRETDEEPEEPGRRGSF 60
          |||
Db      1 MAQALPWL...LWMGAGVLP...AHGTQHGIRLPLR...SGLGGAPLGLRLPRETDEEPEEPGRRGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQT...LNILVDTGSSNFAVGAAPHPFLHRYYQRLSST 120
          |||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQT...LNILVDTGSSNFAVGAAPHPFLHRYYQRLSST 120

Qy    121 YRDLRKGVYVPYTQGWEGELGTDLVSIPHGP...NVTVRANIAAITESDKFFINGSNWEGIL 180
          |||
Db    121 YRDLRKGVYVPYTQGWEGELGTDLVSIPHGP...NVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPD...SLEPFFDSL...VKQTHVP...NLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
          |||
Db    181 GLAYAEIARPD...SLEPFFDSL...VKQTHVP...NLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWY...YEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR...LPKK 300
          |||
Db    241 DHSLYTGSLWYTPIRREWY...YEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR...LPKK 300

Qy    301 VFEEAAVKS...IKAASSTEKFPDGF...LWGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF...RIT 360
          |||
Db    301 VFEEAAVKS...IKAASSTEKFPDGF...LWGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF...RIT 360

Qy    361 ILPQQYLRPVEDVATSQDDCYKFAISQS...SSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
          |||
Db    361 ILPQQYLRPVEDVATSQDDCYKFAISQS...SSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALF...MLPLCLMVCQW 480
          |||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALF...MLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501
          |||
Db    481 RCLRCLRQQHDDFADDISLLK 501
```

RESULT 2

US-09-548-367D-4

; Sequence 4, Application US/09548367D
; Patent No. 6440698
; GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.

```

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR
AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 29915/6280H
; CURRENT APPLICATION NUMBER: US/09/548,367D
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-548-367D-4

```

```

Query Match          99.7%; Score 2656; DB 4; Length 501;
Best Local Similarity 99.8%; Pred. No. 7.4e-267;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      1 MAQALPWLLWLGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRGGSF 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAQALPWLLWLGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRGGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQT LNILVD TGSSNFAVGAAPHPFLHRYYQRQLSST 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQT LNILVD TGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy    121 YRDLRKG VYVPYTQGWEGELGTD LVSIPHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 YRDLRKG VYVPYTQGWEGELGTD LVSIPHGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPD DSLEPF FDSL VKQTHVPNL FSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GLAYAEIARPD DSLEPF FDSL VKQTHVPNL FSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWYVEVI IVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DHSLYTGSLWYTPIRREWYVEVI IVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300

Qy    301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360

Qy    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNI PQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNI PQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

```


Qy 301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRIT 360
 |||||
 Db 301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRIT 360
 Qy 361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
 |||||
 Db 361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
 Qy 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 |||||
 Db 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 Qy 481 RCLRCLRQQHDDFADDISLLK 501
 |||||
 Db 481 RCLRCLRQQHDDFADDISLLK 501

RESULT 4

US-09-009-191-2

; Sequence 2, Application US/09009191

; Patent No. 6319689

; GENERAL INFORMATION:

; APPLICANT: POWELL, DAVID

; APPLICANT: CHAPMAN, CONRAD

; APPLICANT: MURPHY, KAY

; APPLICANT: SMITH, TRUDI

; TITLE OF INVENTION: ASP2

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: RATNER & PRESTIA

; STREET: P.O. BOX 980

; CITY: VALLEY FORGE

; STATE: PA

; COUNTRY: USA

; ZIP: 19482

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/009,191

; FILING DATE: 20-JAN-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: UK 9701684.4

; FILING DATE: 28-JAN-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: PRESTIA, PAUL F

; REGISTRATION NUMBER: 23,031

; REFERENCE/DOCKET NUMBER: GH-70368

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 610-407-0700

; TELEFAX: 610-407-0701

; TELEX: 846169

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 501 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-009-191-2

Query Match 99.5%; Score 2650; DB 4; Length 501;
Best Local Similarity 99.6%; Pred. No. 3.1e-266;
Matches 499; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLPRLSGLGGAPLGLRLPRETDEEPEEPGRGGSF 60
          |||
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLPRLSGLGGAPLGLRLPRETDEEPEEPGRGGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
          |||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy    121 YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
          |||
Db    121 YRDLRKG VYEPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPDDSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
          |||
Db    181 GLAYAEIARPDDSLEPFFDSL VKQTHVPNLFSLQLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
          |||
Db    241 DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300

Qy    301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360
          |||
Db    301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360

Qy    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
          |||
Db    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
          |||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501
          |||
Db    481 RCLRCLRQQHDDFADDISLLK 501
```

RESULT 5

US-09-604-608-2

; Sequence 2, Application US/09604608

; Patent No. 6545127

; GENERAL INFORMATION:

; APPLICANT: Tang, Jordan J.N.

; APPLICANT: Lin, Xinli

; APPLICANT: Koelsch, Gerald

```

; TITLE OF INVENTION: Catalytically Active Recombinant Memapsin and Methods
; TITLE OF INVENTION: of Use Thereof
; FILE REFERENCE: OMRF 179
; CURRENT APPLICATION NUMBER: US/09/604,608
; CURRENT FILING DATE: 2000-06-27
; PRIOR APPLICATION NUMBER: 60/141,363
; PRIOR FILING DATE: 1999-06-28
; PRIOR APPLICATION NUMBER: 60/168,060
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: 60/177,836
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: 60/178,368
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/210,292
; PRIOR FILING DATE: 2000-06-08
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Purified Memapsin 2
; OTHER INFORMATION: Amino Acids 28-48 are remnant putative propeptide
; OTHER INFORMATION: residues
; OTHER INFORMATION: Amino Acids 58-61, 78, 80, 82-83, 116, 118-121,
; OTHER INFORMATION: 156, 166, 174, 246, 274, 276, 278-281, 283, and
; OTHER INFORMATION: 376-377 are residues in contact with the OM99-2
; OTHER INFORMATION: inhibitor
; OTHER INFORMATION: Amino acids 54-57, 61-68, 73-80, 86-89, 109-111,
; OTHER INFORMATION: 113-118, 123-134, 143-154, 165-168, 198-202, and
; OTHER INFORMATION: 220-224 are N-lobe Beta Strands
; OTHER INFORMATION: Amino Acids 184-191 and 210-217 are N-lobe Helices
; OTHER INFORMATION: Amino acids 237-240, 247-249, 251-256, 259-260,
; OTHER INFORMATION: 273-275, 282-285, 316-318, 331-336, 342-348,
; OTHER INFORMATION: 354-357, 366-370, 372-375, 380-383, 390-395,
; OTHER INFORMATION: 400-405, and 418-420 are C-lobe Beta Strands
; OTHER INFORMATION: Amino Acids 286-299, 307-310, 350-353, 384-387,
; OTHER INFORMATION: and 427-431 are C-lobe Helices
US-09-604-608-2

```

```

Query Match          96.9%; Score 2582; DB 4; Length 488;
Best Local Similarity 99.8%; Pred. No. 3.4e-259;
Matches 487; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy      14 AGVLP AHGTQH GIRLPLR SGLGGAPLGLRLPRETDEEPEEPGRRGSFVEMVDNLRGKSGQ 73
          |||
Db      1  AGVLP AHGTQH GIRLPLR SGLGGAPLGLRLPRETDEEPEEPGRRGSFVEMVDNLRGKSGQ 60

Qy      74 GYYVEMTVGSP PQTLN ILVD TGSSNFAVGAAPHPFLH RYYQRQLSSTYRDLRKG VYVPYT 133
          |||
Db      61 GYYVEMTVGSP PQTLN ILVD TGSSNFAVGAAPHPFLH RYYQRQLSSTYRDLRKG VYVPYT 120

Qy     134 QGKWE GELGTD LVSIPHGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPD DS 193
          |||
Db     121 QGKWE GELGTD LVSIPHGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPD DS 180

```

Qy 194 LEFFFDLVKQTHVFNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSLWYTP 253
 |||
 Db 181 LEFFFDLVKQTHVFNLFSLQLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSLWYTP 240
 Qy 254 IRREWYVEVIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPKVFEEAAVKSIIKAAS 313
 |||
 Db 241 IRREWYVEVIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPKVFEEAAVKSIIKAAS 300
 Qy 314 STEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRTILPQQYLRPVEDV 373
 |||
 Db 301 STEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRTILPQQYLRPVEDV 360
 Qy 374 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSACHVHDEFRTAAVEG 433
 |||
 Db 361 ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSACHVHDEFRTAAVEG 420
 Qy 434 PFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQWRCLRCLRQQHDDF 493
 |||
 Db 421 PFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQWRCLRCLRQQHDDF 480
 Qy 494 ADDISLLK 501
 |||
 Db 481 ADDISLLK 488

RESULT 6

US-09-713-158-2

; Sequence 2, Application US/09713158

; Patent No. 6361975

; GENERAL INFORMATION:

; APPLICANT: ZHU, YUAN

; APPLICANT: LI, XIAOTONG

; APPLICANT: POWELL, DAVID J.

; APPLICANT: CHRISTIE, GARY

; TITLE OF INVENTION: MOUSE ASPARTIC SECRETASE-2 (MASP-2)

; FILE REFERENCE: GP-70660

; CURRENT APPLICATION NUMBER: US/09/713,158

; CURRENT FILING DATE: 2000-11-15

; PRIOR APPLICATION NUMBER: 60/165,800

; PRIOR FILING DATE: 1999-11-16

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: FastSEQ for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 501

; TYPE: PRT

; ORGANISM: MUS MUSCULUS

US-09-713-158-2

Query Match 96.9%; Score 2582; DB 4; Length 501;

Best Local Similarity 96.6%; Pred. No. 3.6e-259;

Matches 484; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

Qy 1 MAQALPWLLWVGAGVLPAGHTQHGIPLRSLGGLPLGLRLPRETDEEPEEPGRGGSF 60
 |||:|:| || |||
 Db 1 MAQALPWLLWVGSGMLPAQGTHLGIPLRSLGGLPLGLRLPRETDEESEPGRGGSF 60
 Qy 61 VEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFAVGAAPHPFLHRYQRQLSST 120

Db	61	VEMVDNLRGKSGQGYVEMTVGSPQTNLNVLDTGSSNFAVGAAPHPFLHRYYQRLSST	120
Qy	121	YRDLRKGVYVPYTQGWEGELGTDLVSIHPGNVTVRANIAAITESDKFFINGSNWEGIL	180
Db	121	YRDLRKGVYVPYTQGWEGELGTDLVSIHPGNVTVRANIAAITESDKFFINGSNWEGIL	180
Qy	181	GLAYAEIARPDDSLEPFFDSLQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI	240
Db	181	GLAYAEIARPDDSLEPFFDSLQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI	240
Qy	241	DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK	300
Db	241	DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK	300
Qy	301	VFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT	360
Db	301	VFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT	360
Qy	361	ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC	420
Db	361	ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC	420
Qy	421	HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW	480
Db	421	HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW	480
Qy	481	RCLRCLRQHHDDFADDISLLK	501
Db	481	RCLRCLRQHHDDFADDISLLK	501

RESULT 7

US-09-604-608-3

; Sequence 3, Application US/09604608

; Patent No. 6545127

; GENERAL INFORMATION:

; APPLICANT: Tang, Jordan J.N.

; APPLICANT: Lin, Xinli

; APPLICANT: Koelsch, Gerald

; TITLE OF INVENTION: Catalytically Active Recombinant Memapsin and Methods

; TITLE OF INVENTION: of Use Thereof

; FILE REFERENCE: OMRF 179

; CURRENT APPLICATION NUMBER: US/09/604,608

; CURRENT FILING DATE: 2000-06-27

; PRIOR APPLICATION NUMBER: 60/141,363

; PRIOR FILING DATE: 1999-06-28

; PRIOR APPLICATION NUMBER: 60/168,060

; PRIOR FILING DATE: 1999-11-30

; PRIOR APPLICATION NUMBER: 60/177,836

; PRIOR FILING DATE: 2000-01-25

; PRIOR APPLICATION NUMBER: 60/178,368

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: 60/210,292

; PRIOR FILING DATE: 2000-06-08

; NUMBER OF SEQ ID NOS: 31

; SOFTWARE: PatentIn Ver. 2.1

```
; SEQ ID NO 3
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Pro-memapsin 2
; OTHER INFORMATION: Amino Acids 1-15 are vector-derived residues
; OTHER INFORMATION: Amino Acids 16-64 are a putative pro peptide
; OTHER INFORMATION: Amino Acids 1-13 are the T7 promoter
; OTHER INFORMATION: Amino Acids 16-456 are Pro-memapsin 2-T1
; OTHER INFORMATION: Amino Acids 16-421 are Promemapsin 2-T2
US-09-604-608-3
```

```
Query Match          96.9%; Score 2582; DB 4; Length 503;
Best Local Similarity 99.8%; Pred. No. 3.6e-259;
Matches 487; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      14 AGVLPAHGTQHGI R L P L R S G L G G A P L G L R L P R E T D E E P E E P G R R G S F V E M V D N L R G K S G Q 73
          |||
Db      16 AGVLPAHGTQHGI R L P L R S G L G G A P L G L R L P R E T D E E P E E P G R R G S F V E M V D N L R G K S G Q 75

Qy      74 GYYVEMTVGSP PQ T L N I L V D T G S S N F A V G A A P H P F L H R Y Y Q R Q L S S T Y R D L R K G V Y V P Y T 133
          |||
Db      76 GYYVEMTVGSP PQ T L N I L V D T G S S N F A V G A A P H P F L H R Y Y Q R Q L S S T Y R D L R K G V Y V P Y T 135

Qy     134 QGKWE G E L G T D L V S I P H G P N V T V R A N I A A I T E S D K F F I N G S N W E G I L G L A Y A E I A R P D D S 193
          |||
Db     136 QGKWE G E L G T D L V S I P H G P N V T V R A N I A A I T E S D K F F I N G S N W E G I L G L A Y A E I A R P D D S 195

Qy     194 L E P F F D S L V K Q T H V P N L F S L H L C G A G F P L N Q S E V L A S V G G S M I I G G I D H S L Y T G S L W Y T P 253
          |||
Db     196 L E P F F D S L V K Q T H V P N L F S L Q L C G A G F P L N Q S E V L A S V G G S M I I G G I D H S L Y T G S L W Y T P 255

Qy     254 I R R E W Y Y E V I I V R V E I N G Q D L K M D C K E Y N Y D K S I V D S G T T N L R L P K K V F E A A V K S I K A A S 313
          |||
Db     256 I R R E W Y Y E V I I V R V E I N G Q D L K M D C K E Y N Y D K S I V D S G T T N L R L P K K V F E A A V K S I K A A S 315

Qy     314 S T E K F P D G F W L G E Q L V C W Q A G T T P W N I F P V I S L Y L M G E V T N Q S F R I T I L P Q Q Y L R P V E D V 373
          |||
Db     316 S T E K F P D G F W L G E Q L V C W Q A G T T P W N I F P V I S L Y L M G E V T N Q S F R I T I L P Q Q Y L R P V E D V 375

Qy     374 A T S Q D D C Y K F A I S Q S S T G T V M G A V I M E G F Y V V F D R A R K R I G F A V S A C H V H D E F R T A A V E G 433
          |||
Db     376 A T S Q D D C Y K F A I S Q S S T G T V M G A V I M E G F Y V V F D R A R K R I G F A V S A C H V H D E F R T A A V E G 435

Qy     434 P F V T L D M E D C G Y N I P Q T D E S T L M T I A Y V M A A I C A L F M L P L C L M V C Q W R C L R C L R Q Q H D D F 493
          |||
Db     436 P F V T L D M E D C G Y N I P Q T D E S T L M T I A Y V M A A I C A L F M L P L C L M V C Q W R C L R C L R Q Q H D D F 495

Qy     494 A D D I S L L K 501
          |||
Db     496 A D D I S L L K 503
```

```
RESULT 8
US-09-548-372D-8
; Sequence 8, Application US/09548372D
```

```

; Patent No. 6420534
; GENERAL INFORMATION:
; APPLICANT: GURNEY ET AL.
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR
AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 29915/6280I
; CURRENT APPLICATION NUMBER: US/09/548,372D
; CURRENT FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-548-372D-8

```

```

Query Match          96.4%; Score 2567; DB 4; Length 501;
Best Local Similarity 96.2%; Pred. No. 1.3e-257;
Matches 482; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

```

```

Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRRGSF 60
        || || |||||:|:|:| || ||||| || ||||| ||||| |||||
Db      1 MAPALHWLLLWVGSGMLPAQGTHLGI RLP LRSGLAGPPLGLRLPRETDEESEEPGRRGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPPQTLN I LVD TGSSNFAVGAAPHPFLHRYYQRQLSST 120
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPPQTLN I LVD TGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy    121 YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    121 YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPDDSLEPFFDSL VKQTHV PNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    181 GLAYAEIARPDDSLEPFFDSL VKQTHI PNI FSLQLCGAGFPLNQTEALASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWY YEV I VRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    241 DHSLYTGSLWYTPIRREWY YEV I VRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300

Qy    301 VFEEAVKSIKAASSTEKFPDGF LGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    301 VFEEAVKSIKAASSTEKFPDGF LGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360

Qy    361 ILPQQYL R PVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db    361 ILPQQYL R PVEDVATSQDDCYKFAVSQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

```

QY 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 |||||
 Db 421 HVHDEFRTAAVEGPFVTADMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 QY 481 RCLRCLRQOHDDFADDISLLK 501
 |||||
 Db 481 RCLRCLRQOHDDFADDISLLK 501

RESULT 9

US-09-548-367D-8

; Sequence 8, Application US/09548367D

; Patent No. 6440698

; GENERAL INFORMATION:

; APPLICANT: GURNEY ET AL.

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: 29915/6280H

; CURRENT APPLICATION NUMBER: US/09/548,367D

; CURRENT FILING DATE: 2000-04-12

; PRIOR APPLICATION NUMBER: US 60/155,493

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 09/404,133

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: PCT/US99/20881

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 60/101,594

; PRIOR FILING DATE: 1998-09-24

; NUMBER OF SEQ ID NOS: 73

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 8

; LENGTH: 501

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-548-367D-8

Query Match 96.4%; Score 2567; DB 4; Length 501;

Best Local Similarity 96.2%; Pred. No. 1.3e-257;

Matches 482; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

QY 1 MAQALPWLILLWMGAGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEEPEEPGRGGSF 60
 || || |||||:|:|:| || ||||| | ||||| |||||
 Db 1 MAPALHWLLLWVGSGMLPAQGTHLGIRLPLRSGLAGPPLGLRLPRETDEESEEPGRGGSF 60
 QY 61 VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
 |||||
 Db 61 VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
 QY 121 YRDLRKGVPYPYTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180
 |||||
 Db 121 YRDLRKGVPYPYTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180
 QY 181 GLAYAEIARPDSDLPPFFDSLQKTHVFNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
 |||||:|:|:| |||||:| |||||
 Db 181 GLAYAEIARPDSDLPPFFDSLQKTHIPNIFSLQCGAGFPLNQTEALASVGGSMIIGGI 240

Qy 241 DHSLYTGSLWYTPIRREWYVEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPKK 300
 |||
 Db 241 DHSLYTGSLWYTPIRREWYVEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPKK 300

Qy 301 VFEEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
 |||
 Db 301 VFEEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360

Qy 361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
 |||:|
 Db 361 ILPQQYLRPVEDVATSQDDCYKFAVSQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 |||
 Db 421 HVHDEFRTAAVEGPFVTADMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy 481 RCLRCLRQHHDDFADDISLLK 501
 |||
 Db 481 RCLRCLRHHQHHDDFADDISLLK 501

RESULT 10

US-09-551-853D-8

; Sequence 8, Application US/09551853D

; Patent No. 6500667

; GENERAL INFORMATION:

; APPLICANT: GURNEY ET AL.

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR
 AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: 29915/6280L

; CURRENT APPLICATION NUMBER: US/09/551,853D

; CURRENT FILING DATE: 2000-04-18

; PRIOR APPLICATION NUMBER: US 60/155,493

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 09/404,133

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: PCT/US99/20881

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 60/101,594

; PRIOR FILING DATE: 1998-09-24

; NUMBER OF SEQ ID NOS: 73

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 8

; LENGTH: 501

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-551-853D-8

Query Match 96.4%; Score 2567; DB 4; Length 501;

Best Local Similarity 96.2%; Pred. No. 1.3e-257;

Matches 482; Conservative 7; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MAQALPWLLLWMGAGVLPAGHTQHGIPLRLSGLGGAPLGLRLPRETDEEPEEPGRGSF 60
 || || |||:|:| || |||
 Db 1 MAPALHWLLLWVGSGMLPAQGTHLGIPLRLSGLAGPPLGLRLPRETDEESEEPEEPGRGSF 60

Qy	61	VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHFPLHRYYQRQLSST	120
Db	61	VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHFPLHRYYQRQLSST	120
Qy	121	YRDLRKGVPYPTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL	180
Db	121	YRDLRKGVPYPTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL	180
Qy	181	GLAYAEIARPPDSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI	240
		: :	
Db	181	GLAYAEIARPPDSLEPFFDSL VKQTHIPNIFSLQLCGAGFPLNQTEALASVGGSMIIGGI	240
Qy	241	DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRLPKK	300
Db	241	DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRLPKK	300
Qy	301	VFEAAVKSIIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT	360
Db	301	VFEAAVKSIIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT	360
Qy	361	ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC	420
		:	
Db	361	ILPQQYLRPVEDVATSQDDCYKFAVSQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC	420
Qy	421	HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW	480
Db	421	HVHDEFRTAAVEGPFVTADMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW	480
Qy	481	RCLRCLRQHHDDFADDISLLK	501
Db	481	RCLRCLRHHQHHDDFADDISLLK	501

RESULT 11

US-09-548-372D-6

; Sequence 6, Application US/09548372D

; Patent No. 6420534

; GENERAL INFORMATION:

; APPLICANT: GURNEY ET AL.

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: 29915/6280I

; CURRENT APPLICATION NUMBER: US/09/548,372D

; CURRENT FILING DATE: 2000-04-12

; PRIOR APPLICATION NUMBER: US 60/155,493

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 09/404,133

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: PCT/US99/20881

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 60/101,594

; PRIOR FILING DATE: 1998-09-24

; NUMBER OF SEQ ID NOS: 73

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 6

; LENGTH: 476

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-548-372D-6

Query Match 94.1%; Score 2506.5; DB 4; Length 476;
Best Local Similarity 95.0%; Pred. No. 2.3e-251;
Matches 476; Conservative 0; Mismatches 0; Indels 25; Gaps 1;

```
Qy      1 MAQALPWL...LWMGAGVLP...AHGTQH...GIRLPLR...SGLGGAPL...GLRLP...RETDEEPEEPGRG...SF 60
          |||
Db      1 MAQALPWL...LWMGAGVLP...AHGTQH...GIRLPLR...SGLGGAPL...GLRLP...RETDEEPEEPGRG...SF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSP...QTLN...ILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
          |||
Db     61 VEMVDNLRGKSGQGYVEMTVGSP...QTLN...ILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy    121 YRDLRKG...VYPYTQ...GKWE...GELGTD...LVSI...PHGPNVTVRANIAA...ITESDKFFINGSN...WEGIL 180
          |||
Db    121 YRDLRKG...VYPYTQ...GKWE...GELGTD...LVSI...PHGPNVTVRANIAA...ITESDKFFINGSN...WEGIL 180

Qy    181 GLAYAEIAR...PDDSL...EPFFD...SLVKQTHV...PNL...FSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
          |||
Db    181 GLAYAEIAR-----LCGAGFPLNQSEVLASVGGSMIIGGI 215

Qy    241 DHSLYTGSLWYTPIRREWY...YEVII...VRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR...LPKK 300
          |||
Db    216 DHSLYTGSLWYTPIRREWY...YEVII...VRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR...LPKK 275

Qy    301 VFEEAAVKS...IKAASSTEKFPD...GFWLGEQLVCWQAGTTPWNI...FPVISLYLMGEVTNQSF...RIT 360
          |||
Db    276 VFEEAAVKS...IKAASSTEKFPD...GFWLGEQLVCWQAGTTPWNI...FPVISLYLMGEVTNQSF...RIT 335

Qy    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
          |||
Db    336 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 395

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNI...PQTDESTLMTIAYVMAAICALF...MLPLCLMVCQW 480
          |||
Db    396 HVHDEFRTAAVEGPFVTLDMEDCGYNI...PQTDESTLMTIAYVMAAICALF...MLPLCLMVCQW 455

Qy    481 RCLRCLRQQHDDFADDISLLK 501
          |||
Db    456 RCLRCLRQQHDDFADDISLLK 476
```

RESULT 12

US-09-548-367D-6

; Sequence 6, Application US/09548367D

; Patent No. 6440698

; GENERAL INFORMATION:

; APPLICANT: GURNEY ET AL.

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR
AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: 29915/6280H

; CURRENT APPLICATION NUMBER: US/09/548,367D

; CURRENT FILING DATE: 2000-04-12

```
; PRIOR APPLICATION NUMBER: US 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: US 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 476
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-548-367D-6
```

```
Query Match          94.1%; Score 2506.5; DB 4; Length 476;
Best Local Similarity 95.0%; Pred. No. 2.3e-251;
Matches 476; Conservative 0; Mismatches 0; Indels 25; Gaps 1;
```

```
Qy      1 MAQALPWL LLWMGAGV LPAHGTQH GIRLPLR SGLGGAP LGLRLP RETDEEPEEPGR RGSF 60
      |||||||
Db      1 MAQALPWL LLWMGAGV LPAHGTQH GIRLPLR SGLGGAP LGLRLP RETDEEPEEPGR RGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSP PQTLN I LVD TGSSNFAVGAAPHPFLH RYYQRQLSST 120
      |||||||
Db     61 VEMVDNLRGKSGQGYVEMTVGSP PQTLN I LVD TGSSNFAVGAAPHPFLH RYYQRQLSST 120

Qy    121 YRDLRKG VYVPYTQ GKWEGELGTD LVSIPHGPNVTVRANIAA ITESDKFFINGSNWEGIL 180
      |||||||
Db    121 YRDLRKG VYVPYTQ GKWEGELGTD LVSIPHGPNVTVRANIAA ITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIAR PDDSLEPFFDSL VKQTHV PNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
      |||||||
Db    181 GLAYAEIAR-----LCGAGFPLNQSEVLASVGGSMIIGGI 215

Qy    241 DHSLYTGSLWYTPIRREWYYEVI IVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LRPKK 300
      |||||||
Db    216 DHSLYTGSLWYTPIRREWYYEVI IVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LRPKK 275

Qy    301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360
      |||||||
Db    276 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 335

Qy    361 ILPQQYL RLPVEDVAT SQDDCYKFAISQSSTGTVMGAVIMEGFYV VFD RARKRIGFAVSAC 420
      |||||||
Db    336 ILPQQYL RLPVEDVAT SQDDCYKFAISQSSTGTVMGAVIMEGFYV VFD RARKRIGFAVSAC 395

Qy    421 HVHDEFRTAAVEGPFVTLD MEDCGYNIPQTDESTLMTIAYVMAA ICALEFMLPLCLMVCQW 480
      |||||||
Db    396 HVHDEFRTAAVEGPFVTLD MEDCGYNIPQTDESTLMTIAYVMAA ICALEFMLPLCLMVCQW 455

Qy    481 RCLRCLRQQHDDFADDISLLK 501
      |||||||
Db    456 RCLRCLRQQHDDFADDISLLK 476
```

RESULT 13
 US-09-551-853D-6
 ; Sequence 6, Application US/09551853D
 ; Patent No. 6500667
 ; GENERAL INFORMATION:
 ; APPLICANT: GURNEY ET AL.
 ; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR
 AND USES
 ; TITLE OF INVENTION: THEREOF
 ; FILE REFERENCE: 29915/6280L
 ; CURRENT APPLICATION NUMBER: US/09/551,853D
 ; CURRENT FILING DATE: 2000-04-18
 ; PRIOR APPLICATION NUMBER: US 60/155,493
 ; PRIOR FILING DATE: 1999-09-23
 ; PRIOR APPLICATION NUMBER: US 09/404,133
 ; PRIOR FILING DATE: 1999-09-23
 ; PRIOR APPLICATION NUMBER: PCT/US99/20881
 ; PRIOR FILING DATE: 1999-09-23
 ; PRIOR APPLICATION NUMBER: US 60/101,594
 ; PRIOR FILING DATE: 1998-09-24
 ; NUMBER OF SEQ ID NOS: 73
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 6
 ; LENGTH: 476
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-551-853D-6

Query Match 94.1%; Score 2506.5; DB 4; Length 476;
 Best Local Similarity 95.0%; Pred. No. 2.3e-251;
 Matches 476; Conservative 0; Mismatches 0; Indels 25; Gaps 1;

Qy	1	MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGGSF	60
Db	1	MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGGSF	60
Qy	61	VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAHPFLHRYYQRQLSST	120
Db	61	VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAHPFLHRYYQRQLSST	120
Qy	121	YRDLRKGVPYPTQGKWEDELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL	180
Db	121	YRDLRKGVPYPTQGKWEDELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL	180
Qy	181	GLAYAEIARPDSDLVQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI	240
Db	181	GLAYAEIAR-----LCGAGFPLNQSEVLASVGGSMIIGGI	215
Qy	241	DHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDKSIVDSGTTNLRPKK	300
Db	216	DHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNDKSIVDSGTTNLRPKK	275
Qy	301	VFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT	360
Db	276	VFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT	335
Qy	361	ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC	420


```

          |||||
Db      181 GLAYAEIAR-----LCGAGFPLNQTEALASVGGSMIIGGI 215
Qy      241 DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK 300
          |||||
Db      216 DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK 275
Qy      301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
          |||||
Db      276 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 335
Qy      361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
          |||||
Db      336 ILPQQYLRPVEDVATSQDDCYKFAVSQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 395
Qy      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
          |||||
Db      396 HVHDEFRTAAVEGPFVTADMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 455
Qy      481 RCLRCLRQQHDDFADDISLLK 501
          |||||
Db      456 RCLRCLRHQHDDFADDISLLK 476

```

RESULT 15

US-09-548-367D-73

; Sequence 73, Application US/09548367D

; Patent No. 6440698

; GENERAL INFORMATION:

; APPLICANT: GURNEY ET AL.

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR AND USES

; TITLE OF INVENTION: THEREOF

; FILE REFERENCE: 29915/6280H

; CURRENT APPLICATION NUMBER: US/09/548,367D

; CURRENT FILING DATE: 2000-04-12

; PRIOR APPLICATION NUMBER: US 60/155,493

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 09/404,133

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: PCT/US99/20881

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: US 60/101,594

; PRIOR FILING DATE: 1998-09-24

; NUMBER OF SEQ ID NOS: 73

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 73

; LENGTH: 476

; TYPE: PRT

; ORGANISM: Mus musculus

US-09-548-367D-73

Query Match 90.9%; Score 2420.5; DB 4; Length 476;

Best Local Similarity 91.8%; Pred. No. 1.9e-242;

Matches 460; Conservative 5; Mismatches 11; Indels 25; Gaps 1;

Qy 1 MAQALPWLLWLGAGVLPAGHTQHGIPLRLSGLGAPLGLRLPRETDEEPEEPGRRSF 60

Db	1	MAPALHWLLWVGSGMLPAQGTHLGIRLPLRSGLAGPPLGLRLPRETDEESEEPGRRGSF	60
Qy	61	VEMVDNLRGKSGQGYVEMTVGSPPQTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST	120
Db	61	VEMVDNLRGKSGQGYVEMTVGSPPQTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST	120
Qy	121	YRDLRKGVPYPTQGKWEDELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL	180
Db	121	YRDLRKGVPYPTQGKWEDELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL	180
Qy	181	GLAYAEIARPDSDLPPFFDSLKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI	240
Db	181	GLAYAEIAR-----LCGAGFPLNQTEALASVGGSMIIGGI	215
Qy	241	DHSLYTGSLWYTPIRREWYVEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPKK	300
Db	216	DHSLYTGSLWYTPIRREWYVEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPKK	275
Qy	301	VFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT	360
Db	276	VFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT	335
Qy	361	ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC	420
Db	336	ILPQQYLRPVEDVATSQDDCYKFAVSQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC	395
Qy	421	HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW	480
Db	396	HVHDEFRTAAVEGPFVTADMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW	455
Qy	481	RCLRCLRQHQHDDFADDISLLK	501
Db	456	RCLRCLRQHQHDDFADDISLLK	476

Search completed: January 21, 2004, 09:27:07
Job time : 46.0229 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 09:16:55 ; Search time 45.9809 Seconds
(without alignments)
1047.838 Million cell updates/sec

Title: US-09-869-414A-4
Perfect score: 2664
Sequence: 1 MAQALPWLLLWMGAGVLPAAH.....CLRCLRQQHDDFADDISLLK 501

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	2656	99.7	501	2	A59090	aspartic proteinas
2	324	12.2	384	2	JC7574	pepsinogen A - Afr
3	313.5	11.8	385	2	JC7575	pepsinogen A - bul
4	309	11.6	387	2	B38302	pepsin (EC 3.4.23.
5	307.5	11.5	383	2	JC7573	pepsinogen C - Afr
6	307.5	11.5	388	1	S19682	pepsin A (EC 3.4.2
7	305	11.4	384	2	A39314	gastricsin (EC 3.4
8	304	11.4	382	1	PECH	pepsin A (EC 3.4.2
9	301.5	11.3	383	2	A41443	pepsin (EC 3.4.23.
10	301.5	11.3	396	2	A34401	cathepsin E (EC 3.
11	300.5	11.3	412	1	KHHUD	cathepsin D (EC 3.
12	299	11.2	387	2	C38302	pepsin (EC 3.4.23.
13	298	11.2	387	2	D38302	pepsin (EC 3.4.23.

14	297	11.1	407	1	KHRTD	cathepsin D (EC 3.
15	295	11.1	391	2	A43356	cathepsin E (EC 3.
16	294.5	11.1	388	1	S19684	pepsin A (EC 3.4.2
17	290.5	10.9	380	2	I47176	chymosin (EC 3.4.2
18	289	10.8	387	2	E38302	pepsin (EC 3.4.23.
19	288.5	10.8	388	1	PEHU	pepsin A (EC 3.4.2
20	288.5	10.8	388	1	PEMQAR	pepsin A (EC 3.4.2
21	288.5	10.8	388	2	A30142	pepsin A (EC 3.4.2
22	288.5	10.8	398	2	S66465	cathepsin E (EC 3.
23	288	10.8	389	2	JE0371	pepsin C (EC 3.4.2
24	287.5	10.8	444	2	T24204	hypothetical prote
25	287	10.8	410	1	KHMSD	cathepsin D (EC 3.
26	286.5	10.8	388	1	PEMQAJ	pepsin A (EC 3.4.2
27	286	10.7	398	2	I51185	cathepsin D (EC 3.
28	285.5	10.7	388	2	B30142	pepsin A (EC 3.4.2
29	284.5	10.7	381	1	CMSHB	chymosin (EC 3.4.2
30	281.5	10.6	386	1	PEPG	pepsin A (EC 3.4.2
31	281	10.5	387	2	JC7245	pepsinogen A - com
32	276.5	10.4	381	1	CMBO	chymosin (EC 3.4.2
33	276.5	10.4	396	2	S36865	cathepsin E (EC 3.
34	273.5	10.3	376	2	I45856	aspartic proteinas
35	273.5	10.3	377	1	PEMQCJ	gastricsin (EC 3.4
36	270.5	10.2	381	2	JC7247	prochymosin - comm
37	270.5	10.2	389	2	A38302	pepsin (EC 3.4.23.
38	269.5	10.1	344	1	KHPGD	cathepsin D (EC 3.
39	268	10.1	396	2	T47207	aspartic proteinas
40	266.5	10.0	388	2	JC7246	pepsinogen C - com
41	266	10.0	394	2	B43356	gastricsin (EC 3.4
42	266	10.0	405	2	A25379	saccharopepsin (EC
43	264.5	9.9	388	2	A29937	gastricsin (EC 3.4
44	264	9.9	334	2	JC4870	pepsin A (EC 3.4.2
45	264	9.9	380	2	S03433	candidapepsin (EC

ALIGNMENTS

RESULT 1

A59090

aspartic proteinase (EC 3.4.23.-) BACE precursor - human

N;Alternate names: beta-secretase; beta-site APP cleaving enzyme

C;Species: Homo sapiens (man)

C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 11-May-2000

C;Accession: A59090

R;Vassar, R.; Bennett, B.D.; Babu-Khan, S.; Kahn, S.; Mendiaz, E.A.; Denis, P.;

Teplov, D.B.; Ross, S.; Amarante, P.; Loeloff, R.; Luo, Y.; Fisher, S.; Fuller,

J.; Edenson, S.; Lile, J.; Jarosinski, M.A.; Biere, A.L.; Curran, E.; Burgess,

T.; Louis, J.C.; Collins, F.; Treanor, J.; Rogers, G.; Citron, M.

Science 286, 735-741, 1999

A;Title: beta-Secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE.

A;Reference number: A59090; MUID:20002972; PMID:10531052

A;Note: submitted to GenBank, September 1999

A;Accession: A59090

A;Status: not compared with conceptual translation

A;Molecule type: mRNA

A;Residues: 1-501 <VAS>

A;Cross-references: GB:AF190725; NID:g6118538; PIDN:AAF04142.1; PID:g6118539
 C;Genetics:
 A;Gene: BACE
 C;Superfamily: beta-secretase
 C;Keywords: Alzheimer's disease; aspartic proteinase; brain; glycoprotein;
 hydrolase; protein digestion; transmembrane protein; zymogen
 F;1-21/Domain: signal sequence #status predicted <SIG>
 F;22-45/Domain: propeptide #status predicted <PRO>
 F;46-501/Product: acid proteinase BACE #status predicted <MAT>
 F;461-477/Domain: transmembrane #status predicted <TRN>
 F;93,289/Active site: Asp #status predicted
 F;153,172,223,354/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F;330-380/Disulfide bonds: #status predicted

Query Match 99.7%; Score 2656; DB 2; Length 501;
 Best Local Similarity 99.8%; Pred. No. 6.9e-206;
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MAQALPWLLLWMGAGVLPAGHTQHGI	RLPLRSLGGAPLGLRLPRETDEEPEEPGR	RSF	60
Db	1	MAQALPWLLLWMGAGVLPAGHTQHGI	RLPLRSLGGAPLGLRLPRETDEEPEEPGR	RSF	60
Qy	61	VEMVDNLRGKSGQGYVEMTVGSP	QTLNVLVDTGSSNFAVGAAPHF	LHRYYQRQLSST	120
Db	61	VEMVDNLRGKSGQGYVEMTVGSP	QTLNVLVDTGSSNFAVGAAPHF	LHRYYQRQLSST	120
Qy	121	YRDLRKGVYVPYTQGWEGELG	TDLVSI	PHGPNVTVRANIAAITESDKFFINGSN	WEGIL 180
Db	121	YRDLRKGVYVPYTQGWEGELG	TDLVSI	PHGPNVTVRANIAAITESDKFFINGSN	WEGIL 180
Qy	181	GLAYAEIARPDDSLEPF	FDLSVKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMI	IIGGI	240
Db	181	GLAYAEIARPDDSLEPF	FDLSVKQTHVPNLFSLQLCGAGFPLNQSEVLASVGGSMI	IIGGI	240
Qy	241	DHSLYTGSLWYTPIRREWY	YEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR	LPLPKK	300
Db	241	DHSLYTGSLWYTPIRREWY	YEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR	LPLPKK	300
Qy	301	VFEAAVKSIIKAASSTEKFPD	GFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF	FRIT	360
Db	301	VFEAAVKSIIKAASSTEKFPD	GFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF	FRIT	360
Qy	361	ILPQQYLRPVEDVATSQDDCYKFAISQS	SSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC		420
Db	361	ILPQQYLRPVEDVATSQDDCYKFAISQS	SSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC		420
Qy	421	HVHDEFRTAAVEGPFVTLDMEDCGYNIPQ	TDESTLMTIAYVMAAICALFMLPLCLMVCQW		480
Db	421	HVHDEFRTAAVEGPFVTLDMEDCGYNIPQ	TDESTLMTIAYVMAAICALFMLPLCLMVCQW		480
Qy	481	RCLRCLRQQHDDFADDISLLK			501
Db	481	RCLRCLRQQHDDFADDISLLK			501

RESULT 2
 JC7574

pepsinogen A - African clawed frog
 C;Species: *Xenopus laevis* (African clawed frog)
 C;Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001
 C;Accession: JC7574; PC7119
 R;Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.
 J. Biochem. 129, 147-153, 2001
 A;Title: Amphibian pepsinogens: Purification and characterization of *Xenopus* pepsinogens, and molecular cloning of *Xenopus* and bullfrog pepsinogens.
 A;Reference number: JC7573; MUID:21064922; PMID:11134969
 A;Contents: Stomach
 A;Accession: JC7574
 A;Molecule type: mRNA
 A;Residues: 1-384 <IKU>
 A;Cross-references: DDBJ:AB045380
 A;Accession: PC7119
 A;Molecule type: protein
 A;Residues: 16-35;57-76 <IK2>
 C;Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like activity.
 C;Genetics:
 A;Gene: PgA
 C;Superfamily: pepsin
 C;Keywords: stomach; zymogen

Query Match 12.2%; Score 324; DB 2; Length 384;
 Best Local Similarity 25.5%; Pred. No. 3.1e-18;
 Matches 113; Conservative 73; Mismatches 158; Indels 100; Gaps 19;

```

Qy      8 LLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRRGSFVE----- 62
      ||| :| || :: ::||| | | : | | ::
Db      3 LLLLLGLVVL----SECVKVPLRKG-----ESFRNRPQRLGLLDYLLKKNPYN 47

Qy     63 -----MVDNLRGKSGQGYVEMTVGSPQTLNIIIVDTGSSNFAVGAAPHPFL- 109
      : | : || ::|:| | : |||:| | | :
Db     48 PASKYFPTLAQSSAETLQNYMDIEYYGTISIGTPPQEFTVIFDTGSANLWV---PSVYCS 104

Qy    110 -----HRYYQRQLSSTYRDLRKGVYVPYTQGWEGELGTDLVSIHPGPNVTVRANIAAI 163
      | : | |||: | : | | || | : : | : :
Db    105 SQACSNHNRFNPQQSSTFQATNTPVSIQYGTGSMGFLGYDTLQV---GNIQISNQMFGL 161

Qy    164 TESDK-FFINGSNWEGLGLAYAEIARPDDSLEPFDSLQKQTHVP-NLFSLHLCGAGFP 221
      :||: | : | :||| | : || | |||:| | |
Db    162 SESEPGSFLYYSPFDGILGLAFPSIA--SSQATPVFDNMWSQGLIPQNLFSVYLSSDG-- 217

Qy    222 LNQSEVLASVGGSMIIGGIDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDL--KMDCK 279
      | : : |||:| | ||| : | : | : : | ||| : |
Db    218 -----QTGSYVLFGGVDNSYYSGLNWWPLTAETYWQITLDSVVSINGQVIACSQSC- 268

Qy    280 EYNYDKSIVDSGTTNLRLPKKVFEEAVKSIKAASSTEKFPDGFWLGEQLV-CWQAGTTPW 338
      :|||:| : | | : : || | : : | |
Db    269 -----QAIVDTGTSLMTGPSTPI-ANIQNYIGASQDSN-----GQYVINCNNISNMPT 315

Qy    339 NIFPVISLYLMGEVTNQSFRTILPQQY-LRPVEDVATSQDDCYK-FAISQSSTGT---- 392
      :| | | | | | : | | | :
Db    316 IVF-----TINGVQYPLSPSAYVRQNQQGCSSGFQAMNLPNSGDLW 357

Qy    393 VMGAVIMEGFYVVFDRARKRIGFA 416
  
```

Db ::||| : :: ||||| : |
358 ILGDVFIRQYFTVFD RANNYYVAIA 381

RESULT 3

JC7575

pepsinogen A - bullfrog

C;Species: Rana catesbeiana (bullfrog)

C;Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001

C;Accession: JC7575

R;Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.

J. Biochem. 129, 147-153, 2001

A;Title: Amphibian pepsinogens: Purification and characterization of Xenopus pepsinogens, and molecular cloning of Xenopus and bullfrog pepsinogens.

A;Reference number: JC7573; MUID:21064922; PMID:11134969

A;Contents: Stomach

A;Accession: JC7575

A;Molecule type: mRNA

A;Residues: 1-385 <IKU>

A;Cross-references: DDBJ:AB045376

C;Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like activity.

C;Genetics:

A;Gene: PgA

C;Superfamily: pepsin

C;Keywords: stomach; zymogen

Query Match 11.8%; Score 313.5; DB 2; Length 385;
Best Local Similarity 26.6%; Pred. No. 2.2e-17;
Matches 117; Conservative 74; Mismatches 158; Indels 91; Gaps 20;

```
Qy      8 LLLWMGAGVLP AHGTQH GIRLPLRSG LGGAPLGLRLPRETDEEPEEPGRRGSFV----- 61
      :|| | || | ::: || | | || | | | :
Db      3 ILLLFGLVVLAE CGV---VKVSLRK---GESLRARLNR-----LGLLGDY LKKHHYN 48

Qy     62 ---EMVDNLRGKSGQ-----GYVEMTVGSP PQTLN ILVDTGSSNFAVG----AAPH 106
      : :| ||: | : ::|||: : ||||| | : :|
Db     49 PATKYFPSLAQASGEPLQ NYMDIEYFGTISIGT PPQSFTVIFDTGSSNLWVPSVYCSPA 108

Qy    107 PFLHRY YQRQLSSTYRDLRKG VYPYTQ GKWEGELGTDLV SIPHGPNVTVRANIAAITES 166
      | : | |||: | : | | | | | : | : : | : :|
Db    109 CTNHHMFNPQQSSTFQATNTPVSIQYGTGSM SGLGYDTVQV---GNIQITNQIFGLSQS 165

Qy    167 DK-FFINGSNWEGILGLAYAEIARPDDSLEPFFD SLVKQTHVP-NLFSLHLCGAGFPLNQ 224
      : | : | :||| | : | | | : | : | :|||: | | |
Db    166 EPGSFLYYS PFDGILGLAFPSLA--SSQATPVFDNMWNQGLIPQDLFSVYLSSQG----Q 219

Qy    225 SEVLASVGGSMIIGGIDHS LYTGSLWYTPIRREWYYEVIIVRVEINGQDLKM--DCKEYN 282
      | | : ||: | |||: | : | | : : : | | : |
Db    220 S-----GSFVLFGGVDTSY YTG NLNWVPLTAET YWQITVDSISIGGQVIACSGSC---- 269

Qy    283 YDKSIVDSGTTNLR LPKKVFEAAVKS IKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFP 342
      :|||: ||: | | : :| : : : | : : | : |
Db    270 --SAIVDTGTSLLAGP----STPIANI QYYIGANQDSNGQYV---INCNNISNMPTVVF- 319

Qy    343 VISLYLMGEVTNQSF RITILPQQYL RPVED-VATSQDDC---YKFAISQSSTGT--VMGA 396
      || || | | || | : : :| : :|
```

Db 320 -----TINGVQYPLPASAYVRQSQQSCTSGFQAMNLPSSGDLWILGD 362

Qy 397 VIMEGFYVVFDRARKRIGFA 416
 | : : ||||| : |

Db 363 VFIREYYVVFDRANNYVAMA 382

RESULT 4

B38302

pepsin (EC 3.4.23.-) II-1 precursor - rabbit

C;Species: Oryctolagus cuniculus (domestic rabbit)

C;Date: 14-Jun-1991 #sequence_revision 20-Sep-1991 #text_change 23-Feb-1997

C;Accession: B38302

R;Kageyama, T.; Tanabe, K.; Koiwai, O.

J. Biol. Chem. 265, 17031-17038, 1990

A;Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucleotide sequences of cDNAs, molecular evolution, and gene expression during development.

A;Reference number: A38302; MUID:91009127; PMID:2129536

A;Accession: B38302

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-387 <KAG>

A;Cross-references: GB:M59235; GB:J05638

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 11.6%; Score 309; DB 2; Length 387;

Best Local Similarity 27.1%; Pred. No. 5.1e-17;

Matches 98; Conservative 68; Mismatches 130; Indels 66; Gaps 15;

Qy 75 YYVEMTVGSPPQTLNILVDTGSSNFAVG----AAPHFPLHRYYQRQLSSTYRDLRKGVYV 130
 |: :::|:| | :: ||||| | :: |||: : |||: : : :
 Db 75 YFGTISIGTPPQEFTVIFDTGSSNLWVPSTYCSSLACFLHKRFNPDDSSSTFQATSETLSI 134

Qy 131 PYTQGWEGELGTDLVSIHPGNVTVRANIAAITESD---KFFINGSNWEGILGLAYAEI 187
 | | | | | : | : : : : | : : : ||||| |
 Db 135 TYGTGSMTGILGYDTVKV---GNIEDTNQIFGLSKTEPGITFLV--APFDGILGLAYPSI 189

Qy 188 ARPDDSLEPFFDSLQKTHV-PNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYT 246
 : | : | | : : | : |||: | | | : ||| | ||
 Db 190 SASDAT--PVFDNMWNEGLVSEDLFVYLSSNG-----EKGSMVMFGGIDSSYYT 237

Qy 247 GSLWYTPIRREWYVEVIIVRVEINGQDLKM--DCKEYNYDKSIVDSGTTNLRPKKVFEA 304
 ||| : |: | | : : : |||: : | : : |||: | |
 Db 238 GSLNWVPVSHEGYWQITMDSITINGETIACADSC-----QAVVDTGTSLLAGPTSASK 291

Qy 305 AVKSIKAASSTEKFPDGFWLGEQLV-CWQAGTTPWNIFPVISLYLMGEVTNQSFRTILP 363
 | | : : ||| : | : | : | ||
 Db 292 IQSYIGASKNL-----LGENIISCSAIDSLPDIVF-----TINN 325

Qy 364 QQYLRPVED-VATSQDDC---YKFAISQSSTGT--VMGAVIMEGFYVVFDRARKRIGFAV 417
 || | : ||| : : | | : : |||| : | |
 Db 326 VQYPLPASAYILKEDDCLSGFDGMNLDTSYGELWILGDVFIRQYFTVFDRANNQVGLAA 385

Qy 418 SA 419
 : |

RESULT 5

JC7573

pepsinogen C - African clawed frog

N;Alternate names: progastricsin

C;Species: *Xenopus laevis* (African clawed frog)

C;Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 03-Aug-2001

C;Accession: JC7573; PC7118

R;Ikuzawa, M.; Inokuchi, T.; Kobayashi, K.; Yasumasu, S.

J. Biochem. 129, 147-153, 2001

A;Title: Amphibian pepsinogens: Purification and characterization of *Xenopus* pepsinogens, and molecular cloning of *Xenopus* and bullfrog pepsinogens.

A;Reference number: JC7573; MUID:21064922; PMID:11134969

A;Contents: Stomach

A;Accession: JC7573

A;Molecule type: mRNA

A;Residues: 1-383 <IKU>

A;Cross-references: DDBJ:AB045379

A;Accession: PC7118

A;Molecule type: protein

A;Residues: 17-68 <IK2>

C;Comment: This protein is a zymogen for gastric aspartic proteinase, with pepsin-like activity.

C;Genetics:

A;Gene: PgC

C;Superfamily: pepsin

C;Keywords: stomach; zymogen

Query Match 11.5%; Score 307.5; DB 2; Length 383;
Best Local Similarity 25.9%; Pred. No. 6.6e-17;
Matches 112; Conservative 64; Mismatches 139; Indels 117; Gaps 19;

```

Qy      23 QHGIRLPLRSGLGGA PLGLRLPRETDEEPEEPGRRGSFVEMVDNLRGKSGQGYVEMTVG 82
          ::||: ||          | : : : | : | | |||::|
Db      34 ENGIKAPL-----VDPATKYYNQYATAYEPLSNYMDMS---YYGEISIG 74

Qy      83 SPPQTLNILVD TGSSNFAVGA-----APHPFLHRYYQRQLSSTYRDLRKGVVYPYTQ 134
          :||| :| ||||| | :          || : |||| : : |
Db      75 TPPQNFLVLFDTGSSNLWVASTYCQSQACTNHPL----FNPSQSSTYSSNQQFSLQYGT 130

Qy     135 GKWEGELGTDLV SIPHPNVTVRANIAAITESDKFFINGSN-----WEGILGLAYAEIAR 189
          | | || | |:| || : :::: |:| :||| || ||
Db     131 GSLTGILGYD TVTI---QNVAISQQEFG LSETEP----GTN FVYAQFDGILGLAYPSIA- 182

Qy     190 PDDSLEPFFDSL VKQTHVPN-LFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGS 248
          :::| : :| :| | || : || : ||:| : |||
Db     183 -VGGATTVMQGM MQQNLLNQPIFGFYLSG-----QS---SQNGGEVAFGGVDQNYTGTQ 232

Qy     249 LWYTPIRREWY YEVIIVRVEINGQD---LKMDCKEYNYDKSIVDSGTTNLR LPKKVFEAA 305
          ::||: | |::| ||| |          :|||:|:| | | : || :
Db     233 IYWTPVTSETY WQIGIQGFSINGQATGWCSQGC-----QAIVDTGTSLLTAPQSVFSSL 286

Qy     306 VKSIKAASSTEKF PDGFWLGEQLVCWQAGTTPWNI--FPVISLYLMG-----EVTN 354
          ::|| |          |: :| || | || : | |
Db     287 IQSIGAQQDQN-----GQYVVSCS-----NIQNLPTISFTISGVSFPLPPSAYVLQ 332

```

```

Qy      355 QS-----FRITILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVF 407
          ||          | || | :|:          ::| | : :| | :|
Db      333 QSSGYCTIGIMPTYLPSQNGQPL-----WILGDVFLREYYSVYD 371

Qy      408 RARKRIGFAVSA 419
          ::||| :|
Db      372 LGNNQVGFATAA 383

```

RESULT 6

S19682

pepsin A (EC 3.4.23.1) 4 precursor - Japanese macaque

N;Alternate names: pepsinogen A isozyme 4

C;Species: Macaca fuscata (Japanese macaque)

C;Date: 22-Nov-1993 #sequence_revision 19-Oct-1995 #text_change 18-Jun-1999

C;Accession: S19682; S16065

R;Kageyama, T.; Tanabe, K.; Koiwai, O.

Eur. J. Biochem. 202, 205-215, 1991

A;Title: Development-dependent expression of isozymogens of monkey pepsinogens and structural differences between them.

A;Reference number: S19681; MUID:92037645; PMID:1935977

A;Accession: S19682

A;Molecule type: mRNA

A;Residues: 1-388 <KAG>

A;Cross-references: EMBL:X59753; NID:g38070; PIDN:CAA42425.1; PID:g38071

A;Note: parts of sequence, including amino ends of pepsinogen and activation intermediates, confirmed by protein sequencing

C;Comment: This is a minor component of pepsin at all post-partum stages.

C;Comment: Although two-step activation is observed, activation is predominantly a one-step process.

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; gastric juice; hydrolase; phosphoprotein; protein digestion; stomach

F;1-15/Domain: signal sequence #status predicted <SIG>

F;16-388/Product: pepsinogen A 4 #status experimental <PPT>

F;16-62/Domain: activation peptide #status experimental <APT>

F;63-388/Product: pepsin A 4 #status experimental <ENZ>

F;38-39/Cleavage site: Leu-Lys (pepsin) #status experimental

F;62-63/Cleavage site: Leu-Ile (pepsin) #status experimental

F;94,277/Active site: Asp #status predicted

F;107-112,268-272,311-344/Disulfide bonds: #status predicted

F;130/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 11.5%; Score 307.5; DB 1; Length 388;

Best Local Similarity 27.6%; Pred. No. 6.8e-17;

Matches 108; Conservative 65; Mismatches 135; Indels 83; Gaps 17;

```

Qy      44 PRETDEEPEEPGRGRGSFVEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFAVGA 103
          |  ||:| |  :::::  |:  : :|:| |  :: ||||| |
Db      60 PTLIDEQPLE-----NYLDV-----EYFGTIGIGTPAQNFTVVFDTGSSNLWV-- 102

Qy      104 APHPFL-----HRYYQRQLSSTYRDLRGVYVPYTQGWEGELGTDLVSIHPGPNVTV 156
          | :  |  | : | |||| | | : | |  || | | :  ::
Db      103 -PSVYCYSLACMDHNLFPQDSSTYRATSKTVSITYGTGSMTGILGYDTVKV---GGISD 158

Qy      157 RANIAAITESDK-FFINGSNWEGILGLAYAEIARPDDSLEPFFDSLVKQTHV-PNLFSLH 214

```

```

      | :|:: ||: : :||| || | :      | ||:: | | :||::
Db      159 TNQIFGLSETEPGFFLYFAPFDGILGLAYPSIS--SSGATPVFDNIWNQRLVSQDLFSVY 216

Qy      215 LCGAGFPLNQSEVLASVGGSMIIGGIDHSlyTGSlyWYTPIRREWYyEVIIVRVEINGQDL 274
      |      :||      | :| ||| | |||| : |: | |:: : : :||: :
Db      217 LSAD----DQS-----GSVVIFFGIDSSYYTGSlnWVPVSVEGYWQISVDSITMNGKTI 266

Qy      275 --KMDCKEYNYDKSIVDSGTTNLRlPKKVFEAAVKSikaASSTEKFPDGFwLGEQLV-CW 331
      |      :|||:|: | |      | |: :::      || :| |
Db      267 ACAKGC-----QAIVDTGTSLLTGPTSPIANIQSDIGASENSD-----GEMVVSCS 312

Qy      332 QAGTTPWNIFPVISlyLMGEVTNQSFritLPQqY-LRPVEDVATsQDDCYK-----FAI 385
      : | :|      || || | | : || |
Db      313 AISSLPDIVF-----TINGVQYPLPPSAYILQSQGSCTSGFQGMdVP 354

Qy      386 SQSSTGTVMGAVIMEGFYVVFDRARKRIGFA 416
      : :| : :| | : : : |||| : :| |
Db      355 TESGELWILGDVFIRQYFTVFDranNQVGLA 385

```

RESULT 7

A39314

gastricsin (EC 3.4.23.3) precursor - bullfrog

C;Species: *Rana catesbeiana* (bullfrog)

C;Date: 19-Jun-1992 #sequence_revision 19-Jun-1992 #text_change 22-Jun-1999

C;Accession: A39314

R;Yakabe, E.; Tanji, M.; Ichinose, M.; Goto, S.; Miki, K.; Kurokawa, K.; Ito, H.; Kageyama, T.; Takahashi, K.

J. Biol. Chem. 266, 22436-22443, 1991

A;Title: Purification, characterization, and amino acid sequences of pepsinogens and pepsins from the esophageal mucosa of bullfrog (*Rana catesbeiana*).

A;Reference number: A39314; MUID:92042186; PMID:1939266

A;Accession: A39314

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-384 <YAK>

A;Cross-references: GB:M73750; NID:g213687; PIDN:AAA49530.1; PID:g213688

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; hydrolase; protein digestion

```

Query Match          11.4%; Score 305; DB 2; Length 384;
Best Local Similarity 24.5%; Pred. No. 1.1e-16;
Matches 105; Conservative 66; Mismatches 146; Indels 112; Gaps 17;

```

```

Qy      24 HGIRLPLRSGLGGAPlGLRLPRETDEEPEEPGRR--GSFVEMVDNLrGKSGQGYyVEMTV 81
      |||: |:      :| : :| : |      || |::
Db      35 HGIKAPV-----VDPATKYNNFATAFEPLANyMDMSyYGEISI 73

Qy      82 GSPPQTLNlLVDTGSSNFaVGAAPHPFL-----HRYyQRQLsSTYrDLRKGVyVPYTQ 134
      |: ||| :| |||| | | :      | : ||:| : : :|
Db      74 GTPPQNfLVlFDTGSSNLWV---PSTYcQSQACTNHPQFNPSQSSySSNQQQfSLQYGT 130

Qy      135 GKWEGELGTDLVsIPHGPnVTVRANIA-----AITESDKFFINGSNWEGILGLAYAE 186
      | | || | | |      |||      : :| | : : : |||||
Db      131 GSLTGILGYDTVQI-----QNIAISQqEFGLSVTEPGTNfVY-AQFDGILGLAYPS 180

Qy      187 IARPDDSLEPFfDSLvkQTHVPN-LFSLHLcGAGFPLNQSEVLASVGGSMIIGGIDHSly 245

```

```

      || :      ::| :  ||: || |      |      || :  ||:| : |
Db      181 IA--EGGATTVMQGMIIQQNLINQPLFAFYLSGQQNSQN-----GGEVAFGGVDQNY 230

Qy      246 TGSLWYTPIRREWYVEVIIVRVEINGQD---LKMDCKEYNYDKSIVDSGTNLR LPKKVF 302
      :| ::||: | |:: |      :|||      |      : |||:|: | | : ||
Db      231 SGQIYWTPVTSETYWQIGIQGFSVNGQATGWCSQGC-----QGIVDTGTSLLTAPQSVF 284

Qy      303 EAAVKSIIKAASSTEKFPDGFWLGEQLV-CWQAGTTPWNIFPVI-----SLYLMGEVT 353
      : ::|| |      |: | | : | | :      | |:: : :
Db      285 SSLMQSIGAQDQDQ-----GQYAVSCSNIQSLPTISFTISGVSFPLPPSAYVLQONS 336

Qy      354 NQ---SFRITILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRAR 410
      | || | :|:      :|| | : :| | :|
Db      337 GYCTIGIMPTYLPSQNGQPL-----WILGDVFLRQYYSVYDLGN 375

Qy      411 KRIGFAVSA 419
      ::||| :|
Db      376 NQVGFAAAA 384

```

RESULT 8

PECH

pepsin A (EC 3.4.23.1) precursor - chicken

N;Alternate names: pepsinogen A

C;Species: Gallus gallus (chicken)

C;Date: 18-Apr-1984 #sequence_revision 01-Dec-2000 #text_change 01-Dec-2000

C;Accession: JE0370; A00984

R;Sakamoto, N.; Saiga, H.; Yasugi, S.

Biochem. Biophys. Res. Commun. 250, 420-424, 1998

A;Title: Analysis of temporal expression pattern and cis-regulatory sequences of chicken pepsinogen A and C.

A;Reference number: JE0370; MUID:98440813; PMID:9753645

A;Accession: JE0370

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-382 <SAK>

A;Cross-references: GB:AB025281; NID:g4589837; PIDN:BAA76891.1; PID:g4589838

R;Baudys, M.; Kostka, V.

Eur. J. Biochem. 136, 89-99, 1983

A;Title: Covalent structure of chicken pepsinogen.

A;Reference number: A00984; MUID:84004412; PMID:6617663

A;Accession: A00984

A;Molecule type: protein

A;Residues: 16-87,'S',89-382 <BAU>

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; gastric juice; glycoprotein; hydrolase; protein digestion; stomach

F;16-57/Domain: activation peptide #status experimental <APT>

F;58-382/Product: pepsin A #status predicted <MAT>

F;92,275/Active site: Asp #status predicted

F;105-110,266-270,305-338/Disulfide bonds: #status experimental

F;128/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 11.4%; Score 304; DB 1; Length 382;

Best Local Similarity 24.0%; Pred. No. 1.3e-16;

Matches 88; Conservative 69; Mismatches 125; Indels 84; Gaps 13;

```

Qy      75 YYVEMTVGSPQTLNILVDTGSSNFAVGAAPHPFL-----HRYYQRQLSSTYRDLRKG 127
      || ::||| | :: ||||| | : : : ||| :
Db      74 YYGTSISGTPQDFTVIFDTGSSNLWV---PSIYCKSSACSNHKRFDPSSSTYVSTNET 130

Qy      128 VYVPYTQGWEGELGTDLVSIHPGNVTVRANIAAITESDK-FFINGSNWEGILGLAYAE 186
      ||: | | | || | |:: : : | : | ::|:: | |::||| ||:
Db      131 VYIAYGTGMSGILGYDTVAV---SSIDVQNQIFGLSETEPGSFFYYCNFDGILGLAFPS 187

Qy      187 IARPPDSLEPFFDSLQKTHV-PNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLY 245
      | : | ||:: | | :|||::| | | : ||| :
Db      188 IS--SSGATPVFDNMMSQHLVAQDLFSVYLSKDG-----ETGSFVLFGGIDPNYT 235

Qy      246 TGS LWYTPIRREWYVEIIVRVEINGQDLK--MDCKEYNYDKSIVDSGTTNLRPKKVFE 303
      | ::| : | |:: : || : : : | : |||: ||: | :| :
Db      236 TKGIYWVPLSAETYWQITMDRVTVGNKYVACFFTC-----QAIVDTGTSLLVMPQGAYN 289

Qy      304 AAVKSIKAASSTE-----KFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQS 356
      :| : :| | | | | : : : |
Db      290 RIIKDLGVSSDGEISCDISKLPD-----VTFHINGHA---- 322

Qy      357 FRITILPQQYLRPVEDVATSQDDCYKFAISQSSTGT-----VMGAVIMEGFYVVFDRAR 410
      : | : :| | | | : || : ||: ||| |||
Db      323 -----FTLPASAYVLNEDGSCMLGFENMGTPTELGEQWILGDVFIREYYVIFDRAN 373

Qy      411 KRIGFA 416
      ::| :
Db      374 NKVGLS 379

```

RESULT 9

A41443

pepsin (EC 3.4.23.-) precursor, embryonic - chicken

C;Species: Gallus gallus (chicken)

C;Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 21-Jul-2000

C;Accession: A41443

R;Hayashi, K.; Agata, K.; Mochii, M.; Yasugi, S.; Eguchi, G.; Mizuno, T.

J. Biochem. 103, 290-296, 1988

A;Title: Molecular cloning and the nucleotide sequence of cDNA for embryonic chicken pepsinogen: phylogenetic relationship with prochymosin.

A;Reference number: A41443; MUID:88227903; PMID:3131317

A;Accession: A41443

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-383 <HAY>

A;Cross-references: GB:D00215; NID:g2760810; PIDN:BAA00153.1; PID:g222853

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; hydrolase; protein digestion

Query Match 11.3%; Score 301.5; DB 2; Length 383;

Best Local Similarity 25.2%; Pred. No. 2e-16;

Matches 90; Conservative 76; Mismatches 124; Indels 67; Gaps 14;

```

Qy      75 YYVEMTVGSPQTLNILVDTGSSNFAVGA----APHPFLHRYYQRQLSSTYRDLRKG VYV 130
      || ::||| || :: ||||| | : : | : : ||| : : :
Db      76 YYGTSISGTPPDFTVVFDTGSSNLWVPSVSCTSPACQSHQMFNPSQSSTYKSTGQNL SI 135

Qy      131 PYTQGWEGELGTDLVSIHPGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARP 190

```

```

      | | || :| | |:: : : :| :||: :||| | :|
Db      136 HYGTGDMEGTVGCDTVTVASLMDTNQLFGLST-SEPGQFFVY-VKFDGILGLGYPSLAA- 192

Qy      191 DDSLEPFFDSL VKQTHV-PNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSL 249
      | : | ||::| :: : |||::| :| :: ||| | :|||:
Db      193 -DGITPVFDNMVNESLLEQNLF SVYLS-----REPMGSMVVFVGGIDESYFTGSI 240

Qy      250 WYTPIRREWYYEVIIVRVEINGQDL--KMDCKEYNYDKSIVDSGTTNLRPKKVFEAAVK 307
      : | : :|:: : : :| |:: | :||:|:: : |
Db      241 NWIPVSYQGYWQISMSDIIVNKQEIACSSGC-----QAIDTGTSLVAGPASDINDIQS 294

Qy      308 SIKAAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISL----YLMGEVTNQSFRTILP 363
      :: | :| || | | :||: :||: :||:
Db      295 AVGANQNT-----YGEYSV-----NCSHILAMPDVVFVIGGI----- 326

Qy      364 QQYLRPVEDVA----TSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFA 416
      || || :| | | :| :| :| :||| | :| |
Db      327 -QY--PVPALAYTEQNGQGTCMSSFQNSSADLWILGDVFIRVYYSIFDRANNRVGLA 380

```

RESULT 10

A34401

cathepsin E (EC 3.4.23.34) precursor - human

C;Species: Homo sapiens (man)

C;Date: 22-Jun-1990 #sequence_revision 22-Jun-1990 #text_change 22-Jun-1999

C;Accession: A42038; A34401; S35663; S34467; A34643; B34643

R;Azuma, T.; Liu, W.; Vander Laan, D.J.; Bowcock, A.M.; Taggart, R.T.

J. Biol. Chem. 267, 1609-1614, 1992

A;Title: Human gastric cathepsin E gene. Multiple transcripts result from alternative polyadenylation of the primary transcripts of a single gene locus at 1q31-q32.

A;Reference number: A42038; MUID:92112877; PMID:1370478

A;Accession: A42038

A;Molecule type: DNA

A;Residues: 1-396 <AZU>

A;Cross-references: GB:M84424; GB:M82847; NID:g181203; PIDN:AAA52300.1;

PID:g181205

A;Note: sequence extracted from NCBI backbone (NCBIN:75963, NCBIN:75966, NCBIN:75971, NCBIN:75974, NCBIN:75977, NCBIN:75979, NCBIN:75981, NCBIN:75988, NCBIN:75990, NCBIP:75991)

R;Azuma, T.; Pals, G.; Mohandas, T.K.; Couvreur, J.M.; Taggart, R.T.

J. Biol. Chem. 264, 16748-16753, 1989

A;Title: Human gastric cathepsin E. Predicted sequence localization to chromosome 1, and sequence homology with other aspartic proteinases.

A;Reference number: A34401; MUID:89380302; PMID:2674141

A;Accession: A34401

A;Molecule type: mRNA

A;Residues: 1-396 <AZ2>

A;Cross-references: GB:J05036; NID:g181193; PIDN:AAA52130.1; PID:g181194

R;Takeda-Ezaki, M.; Yamamoto, K.

Arch. Biochem. Biophys. 304, 352-358, 1993

A;Title: Isolation and biochemical characterization of procathepsin E from human erythrocyte membranes.

A;Reference number: S35663; MUID:93349047; PMID:8346912

A;Accession: S35663

A;Status: preliminary

A;Molecule type: protein

A;Residues: 20-38;54-76 <TAK>
 R;Hill, J.; Montgomery, D.S.; Kay, J.
 FEBS Lett. 326, 101-104, 1993
 A;Title: Human cathepsin E produced in E. coli.
 A;Reference number: S34467; MUID:93314762; PMID:8325357
 A;Accession: S34467
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 57-60,62-81 <HIL>
 R;Athauda, S.B.P.; Matsuzaki, O.; Kageyama, T.; Takahashi, K.
 Biochem. Biophys. Res. Commun. 168, 878-885, 1990
 A;Title: Structural evidence for two isozymic forms and the carbohydrate attachment site of human gastric cathepsin E.
 A;Reference number: A34643; MUID:90241267; PMID:2334440
 A;Accession: A34643
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 54-58,'XXX',62-64,'M',66-89,'X',91-95 <ATH>
 A;Accession: B34643
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 54-59,'X',61-68 <AT2>
 C;Genetics:
 A;Gene: GDB:CTSE
 A;Cross-references: GDB:119821; OMIM:116890
 A;Map position: 1q31-1q31
 C;Superfamily: pepsin
 C;Keywords: aspartic proteinase; blocked amino end; hydrolase; zymogen
 F;1-17/Domain: signal sequence #status predicted <SIG>
 F;18-53/Domain: activation peptide #status predicted <PRO>
 F;54-396/Product: cathepsin E #status predicted <MAT>
 F;18/Modified site: blocked amino end (Gln) (in mature form) (probably pyrrolidone carboxylic acid) #status experimental
 F;96,281/Active site: Asp #status predicted

Query Match 11.3%; Score 301.5; DB 2; Length 396;
 Best Local Similarity 25.8%; Pred. No. 2.1e-16;
 Matches 100; Conservative 68; Mismatches 144; Indels 75; Gaps 16;

Qy	48	DEEPEEPGRGGSFVEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFAVGA----	103
		: : : : : :: : : :	
Db	63	DQSAKEP-----LINYLD-----MEYFGTISIGSPQNFTVIFDTGSSNLWVPSVYCT	110
Qy	104	APHPFLHRYRQQLSSTYRDLRGVYVPYTQGWEGELGTDLVSIHPGPNVTVRANIAAI	163
		: : : : : : : ::	
Db	111	SPACKTHSRFQPSQSSTYSQPGQSFSIQYGTGSLSGIIGADQVSV-EGLTVVGQQFGESV	169
Qy	164	TESDKFFINGSNWEGILGLAYAEIARPDDSLEPFFDSLQKTHVP-NLFSLHLCGAGFPL	222
		: :: : :: : : :: : ::	
Db	170	TEPGQTFVD-AEFDGILGLGYPSLA--VGGVTPVFDNMMAQNLDLPMFSVYM-----	219
Qy	223	NQSEVLASVGGSMIIGGIDHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYN	282
		: : : : : :: : : :	
Db	220	-SSNPEGGAGSELIFGGYDHSFSGSLNWVPVTKQAYWQIALDNIQVGG--TVMFCSE--	274
Qy	283	YDKSIVDSGTTNLRPLPKKVFEAAVKSIIKAASSTEFKPDGFWLGEQLVCWQAGTTPWNIFP	342
		:: : :: : : : : : :	

Db 275 GCQAIVDTGTSLITGPSDKIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMP 321
 Qy 343 VISLYLMGEVTNQSFRTILPQQYLRPVEDVATSQDDCYKFAISQSSTG----- 391
 :: : | : | : | : | : |
 Db 322 DVTFTING-----VPYTLSPYAY--TLDFVDGMQFC-----SSGFQGLDIHPPAG 365
 Qy 392 --TVMGAVIMEGFYVVFDRARKRIGFA 416
 :: | : || ||| : | |
 Db 366 PLWILGDVFIQFYVFDGRNNRVGLA 392

RESULT 11

KHHUD

cathepsin D (EC 3.4.23.5) precursor [validated] - human

N;Alternate names: preprocathepsin D

C;Species: Homo sapiens (man)

C;Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #text_change 15-Sep-2000

C;Accession: A25771; S30749; PC2066; I59236; I57716

R;Faust, P.L.; Kornfeld, S.; Chirgwin, J.M.

Proc. Natl. Acad. Sci. U.S.A. 82, 4910-4914, 1985

A;Title: Cloning and sequence analysis of cDNA for human cathepsin D.

A;Reference number: A25771; MUID:85270436; PMID:3927292

A;Accession: A25771

A;Molecule type: mRNA

A;Residues: 1-412 <FAU>

A;Cross-references: EMBL:M11233; NID:g181179; PIDN:AAB59529.1; PID:g181180

R;Westley, B.R.; May, F.E.B.

Nucleic Acids Res. 15, 3773-3786, 1987

A;Title: Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive human breast cancer cells.

A;Reference number: S30749; MUID:87231068; PMID:3588310

A;Accession: S30749

A;Molecule type: mRNA

A;Residues: 1-412 <WES>

A;Cross-references: EMBL:X05344; NID:g29677; PIDN:CAA28955.1; PID:g29678

R;May, F.E.B.; Smith, D.J.; Westley, B.R.

Gene 134, 277-282, 1993

A;Title: The human cathepsin D-encoding gene is transcribed from an estrogen-regulated and a constitutive start point.

A;Reference number: PC2066; MUID:94085791; PMID:8262386

A;Accession: PC2066

A;Molecule type: DNA

A;Residues: 1-23 <MAY>

A;Cross-references: GB:L12980; NID:g291930; PIDN:AAA16314.1; PID:g455429

A;Experimental source: MCF-7 cell

R;Cavaillès, V.; Augereau, P.; Rochefort, H.

Proc. Natl. Acad. Sci. U.S.A. 90, 203-207, 1993

A;Title: Cathepsin D gene is controlled by a mixed promoter, and estrogens stimulate only TATA-dependent transcription in breast cancer cells.

A;Reference number: I59236; MUID:93126342; PMID:8419924

A;Accession: I59236

A;Status: translation not shown; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-22 <CAV1>

A;Cross-references: GB:S52557; NID:g263124; PIDN:AAD13868.1; PID:g4261568

R;Augereau, P.; Miralles, F.; Cavaillès, V.; Gaudet, C.; Parker, M.; Rochefort, H.

Mol. Endocrinol. 8, 693-703, 1994

A;Title: Characterization of the proximal estrogen-responsive element of human cathepsin D gene.

A;Reference number: I57716; MUID:95021301; PMID:7935485

A;Accession: I57716

A;Status: translation not shown; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-22 <CAV2>

A;Cross-references: GB:S74689; NID:g786350; PIDN:AAD14156.1; PID:g4261856

R;Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.

submitted to the Brookhaven Protein Data Bank, April 1993

A;Reference number: A51839; PDB:1LYA

A;Contents: annotation; X-ray crystallography, 2.5 angstroms, residues 65-161;170-241

R;Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Erickson, J.W.

submitted to the Brookhaven Protein Data Bank, April 1993

A;Reference number: A51840; PDB:1LYB

A;Contents: annotation; X-ray crystallography, 2.5 angstroms, with inhibitor residues 65-161;170-241

R;Baldwin, E.T.; Bhat, T.N.; Gulnik, S.; Hosur, M.V.; Sowder II, R.C.; Cachau, R.E.; Collins, J.; Silva, A.M.; Erickson, J.W.

Proc. Natl. Acad. Sci. U.S.A. 90, 6796-6800, 1993

A;Title: Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design.

A;Reference number: A48229; MUID:93342076; PMID:8393577

A;Contents: annotation; X-ray crystallography, 2.5 angstroms

C;Comment: Cathepsin D is a ubiquitous lysosomal proteinase.

C;Comment: In addition to the propeptide, residues 163-168 and 411-412 are proteolytically removed. Residues 169 and 170 are also partially removed.

C;Comment: The carbohydrate bound to 134-Asn contains a mannose-6-phosphate that is bound near 267-Lys and the phosphotransferase recognition region.

C;Genetics:

A;Gene: GDB:CTSD

A;Cross-references: GDB:120512; OMIM:116840

A;Map position: 11p15.5-11p15.5

C;Function:

A;Description: limited specificity endopeptidase

A;Pathway: intracellular protein degradation

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; glycoprotein; hydrolase; lysosome; protein degradation

F;1-20/Domain: signal sequence #status predicted <SIG>

F;21-64/Domain: propeptide #status predicted <PRO>

F;65-162,169-410/Product: cathepsin D #status experimental <MAT>

F;267,329-356/Region: phosphotransferase recognition

F;91-160,110-117,286-290,329-366/Disulfide bonds: #status experimental

F;97,295/Active site: Asp #status experimental

F;134,263/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 11.3%; Score 300.5; DB 1; Length 412;

Best Local Similarity 26.9%; Pred. No. 2.7e-16;

Matches 123; Conservative 68; Mismatches 170; Indels 97; Gaps 21;

```
Qy      5 LPWLLWLMGAGVLPAGHTQHGIPLR-----SGLGGAPLGL-----RLP 44
          || | : |   ||           :|||           | :||:   |           :|
Db      7 LPLALCLLAA---PASAL---VRIPLHKFTSIRRTMSEVGGSVEDLIAKGPVSKYSQAVP 60
```

```

Qy      45 RETDEEPEEPGRRGSFVEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFAVGAA 104
      |:      |      |:: |      || |: :|:| |      :: ||| | |      | :
Db      61 AVTE-----GPIPEVLKNYMDAQ---YYGEIGIGTPPQCFTVVFDTGSSNLWVPSI 108

Qy      105 PHPFL-----HRYYQRQLSSTYRDLRKGVPYPTQGWEGELGTDLVSI----- 149
      |      |      |      |||      : |      |      |      | | : |
Db      109 HCKLLDIACWIHHKYNDSKSTYVKNGTSTFDIHYGSGSLSGYLSQDTSVPCQSASSASA 168

Qy      150 HGPNTVVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDSDLFFDSLQTHV-P 208
      |      |      |:      : ::|||:| |      |:      :: | | |:|:| |
Db      169 LGGVKVERQVFGEATKQPGITFIAAKFDGILGMAYPRIS--VNNVLPVFDNLMQQKLVDQ 226

Qy      209 NLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGLWYTPIRREWYVEIIVRVE 268
      |:| | :|      : |      || ::|| |      | || | |      : |: |:| : :| |
Db      227 NIFS FYL-----SRDPDAQPGGELMLGGTDSKYYKGSLSYLVNTRKAYWQVHLDQVE 278

Qy      269 I-NGQDLKMDCKEYNYDKSIVDSGTTNLRPKKVFEAAVKSIIKAASSTEKFPDGFWLGEQ 327
      : :| |      |||      : |||:| | : |      | :| |      :      ||
Db      279 VASGLTL---CKE--GCEAIVDTGTSMLVGPVDEVRELQKAIGAVPLIQ-----GEY 325

Qy      328 LV-CWQAGTTPWNIFPVISLYLMGEVTNQSFRTILPQQYLRPVEDVATSQDDCYKFAIS 386
      :: | : |      | |:| | |      : ::| | : | |      : |      :
Db      326 MIPCEKVST-----LPAILTKLGG----KGYKLS--PEDYTLKVSQAGKTL--CLSGFMG 372

Qy      387 Q-----SSTGTVMGAVIMEGFYVVFDRARKRIGFAVSA 419
      |      :| | : :| ||| |      |:| | :|
Db      373 MDIPPPSGPLWILGDVFIGRYTTFDRDNNRVGFEEAA 410

```

RESULT 12

C38302

pepsin (EC 3.4.23.-) II-2/3 precursor - rabbit

C;Species: *Oryctolagus cuniculus* (domestic rabbit)

C;Date: 14-Jun-1991 #sequence_revision 14-Jun-1991 #text_change 23-Feb-1997

C;Accession: C38302

R;Kageyama, T.; Tanabe, K.; Koiwai, O.

J. Biol. Chem. 265, 17031-17038, 1990

A;Title: Structure and development of rabbit pepsinogens. Stage-specific zymogens, nucleotide sequences of cDNAs, molecular evolution, and gene expression during development.

A;Reference number: A38302; MUID:91009127; PMID:2129536

A;Accession: C38302

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-387 <KAG>

A;Cross-references: GB:J05638

C;Superfamily: pepsin

C;Keywords: aspartic proteinase; hydrolase; phosphoprotein; protein digestion

Query Match 11.2%; Score 299; DB 2; Length 387;

Best Local Similarity 26.9%; Pred. No. 3.3e-16;

Matches 97; Conservative 64; Mismatches 134; Indels 66; Gaps 13;

```

Qy      75 YYVEMTVGSPPTLNILVDTGSSNFAVGAAHPHF-----LHRYYQRQLSSTYRDLRKG 127
      |:      ::|:| |      :: ||| | |      | :      ||: : : |||: :
Db      75 YFGTISIGTPPDFTVIFDTGSSNLWV---PSTYCSSLACALHKRNPEDSSTYQGTSET 131

```

Qy	128	VYVPYTQGKWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEI	187
		: : : : : : :	
Db	132	LSITYGTGSMTGILGYDTVKGSIEDTNQIFGLSKTEPSLTFLF--APFDGILGLAYPSI	189
Qy	188	ARPDDSLEPFDSL VKQTHV--PNLFSLHL CGAGFPLNQSEVLASVGGSMIIGGIDHSLYT	246
		: : : : : : : : : :	
Db	190	SSSDAT--PVFDNMWNEGLVSQDLFSVYLSSDD-----EKGSLVMFGGIDSSYYT	237
Qy	247	GSLWYTPIRREWYYEVIIVRVEINGQDLKM--DCKEYNYDKSIVDSGTTNLR LPKKVFEA	304
		: : : : : : : : : : :	
Db	238	GSLNWVPVSYEGYWQITMDSV SINGETIACADSC-----QAIVDTGTSLLTGP----TS	287
Qy	305	AVKSIKAASSTEKFDPDGFWLGEQLV--CWQAGTTPWNIFPVISLYLMGEVTNQSFRTILP	363
		: : : : : : : :	
Db	288	AISNIQSYIGASK----NLLGENVISCSAIDSLPDIVF-----TING	325
Qy	364	QQYLRPVEDVATSQDDCYKFAISQSSTGT-----VMGAVIMEGFYVVFDRARKRIGFAV	417
		: : : : : : : : : :	
Db	326	IQYPLPASAYILKEDDDCTSGLEGMNVD TYTGELWILGDVFIRQYFTVFD RANNQLGLAA	385
Qy	418	S	418
		:	
Db	386	A	386

Qy 178 GILGLAYAEIARPDDSLPFFDSLVKQTHV-PNLFSLHLCGAGFPLNQSEVLASVGGSMI 236
 ||||| | : | : | ||: : | : ||: | : :
 Db 180 GILGLAYPSISSDAT--PVFDNMWNEGLVSQDLFSVYLSSDD-----EKGSLVM 227

Qy 237 IGGIDHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKM--DCKEYNYDKSIVDSGTTN 294
 |||| | |||| : | : | ||: : | ||: : | : ||: ||:
 Db 228 FGGIDSSYYTGSLNWPVSIEGYWQITMDSVSINGETIACADSC-----QAIVDTGTSL 281

Qy 295 LRLPKKVFEAAVKSIIKAASSTEKFPDGFWLGEQLV-CWQAGTTPWNIFPVISLYLMGEVT 353
 | | : : : : | | | : | : | :
 Db 282 LTGP---TSAISNIQSYIGASK----NLLGENVISCSAIDSLPDIVF----- 321

Qy 354 NQSFRTILPQQYLRPVEDVATSQDDCYKFAISQSSTGT-----VMGAVIMEGFYVVF 407
 || || | : | : : | : : || : : ||
 Db 322 -----TINGIQYPLPASAYILKEDDDCTSGLEGMNVDITYTGELWILGDVFIRQYFTVFD 375

Qy 408 RARKRIGFAVS 418
 || : : | :
 Db 376 RANNQLGLAAA 386

RESULT 14

KHRTD

cathepsin D (EC 3.4.23.5) precursor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 18-Jun-1999

C;Accession: S13111; C31918; JQ1177; PQ0222

R;Birch, N.P.; Loh, Y.P.

Nucleic Acids Res. 18, 6445-6446, 1990

A;Title: Cloning, sequence and expression of rat cathepsin D.

A;Reference number: S13111; MUID:91057150; PMID:2243802

A;Accession: S13111

A;Molecule type: mRNA

A;Residues: 1-407 <BIR>

A;Cross-references: EMBL:X54467; NID:g55881; PIDN:CAA38349.1; PID:g55882

R;Yonezawa, S.; Takahashi, T.; Wang, X.; Wong, R.N.S.; Hartsuck, J.A.; Tang, J.
 J. Biol. Chem. 263, 16504-16511, 1988

A;Title: Structures at the proteolytic processing region of cathepsin D.

A;Reference number: A92681; MUID:89034127; PMID:3182800

A;Accession: C31918

A;Molecule type: protein

A;Residues: 134-162, 'T', 164-170 <YON>

R;Fujita, H.; Tanaka, Y.; Noguchi, Y.; Kono, A.; Himeno, M.; Kato, K.

Biochem. Biophys. Res. Commun. 179, 190-196, 1991

A;Title: Isolation and sequencing of a cDNA clone encoding rat liver lysosomal
 cathepsin D and the structure of three forms of mature enzymes.

A;Reference number: JQ1177; MUID:91354249; PMID:1883350

A;Accession: JQ1177

A;Molecule type: mRNA

A;Residues: 1-14, 'A', 16-204, 'N', 206-261, 'N', 263-407 <FUJ>

A;Accession: PQ0222

A;Molecule type: protein

A;Residues: 65-74;118-127;165-174 <FU2>

A;Experimental source: liver

C;Comment: Cathepsin D in rat liver lysosome occurs as a mixture of both a
 single chain form and two types of two chain forms.

C;Function:

A;Description: limited specificity endopeptidase
 A;Pathway: intracellular protein degradation
 C;Superfamily: pepsin
 C;Keywords: aspartic proteinase; glycoprotein; hydrolase; lysosome; protein degradation
 F;1-20/Domain: signal sequence #status predicted <SIG>
 F;21-64/Domain: propeptide #status predicted <PRO>
 F;65-407/Product: cathepsin D, 43K single-chain form #status predicted <MAT>
 F;65-164/Product: (or 65-165) cathepsin D 12K light chain #status predicted <MA2>
 F;65-117/Product: cathepsin D 9K light chain #status predicted <MA4>
 F;118-407/Product: cathepsin D 34K heavy chain #status predicted <MA5>
 F;165-407/Product: (or 166-407) cathepsin D 30K heavy chain #status predicted <MA3>
 F;91-160,110-117,281-285,324-361/Disulfide bonds: #status predicted
 F;97,290/Active site: Asp #status predicted
 F;134,258/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 11.1%; Score 297; DB 1; Length 407;
 Best Local Similarity 26.1%; Pred. No. 5.1e-16;
 Matches 118; Conservative 76; Mismatches 170; Indels 88; Gaps 20;

Qy	6	PWLLLWMGAGVLPAGHTQHGI RLPLR-----SGLGGA--PLGLRLPRETDEEPEEP	54
		: : : : : : : : :	
Db	4	PGVLLLI-LGLLDASSSAL-IRIPLRKFTSIRRTMTVEVGGSVEDLILKGPITKYSMQSSP	61
Qy	55	GRRGSFVEMVDNLRGKSGQGYVEMTVGSPQTNLNILDVTGSSNFAVGAAPHPFL-----	109
		: : : : : :	
Db	62	RTKEPVSELLKNYLDAQ---YYGEIGIGTPPQCFTVVFDTGSSNLWVPSIHCKLLDIACW	118
Qy	110	-HRYYQRQLSSTYRDLRGKVYPYTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDK	168
		: : : : : :	
Db	119	VHHKYNSDKSSSTYVKNGTSTFDIHYGSGSLSGYLSQDTSVSP-----CKSDLGGIKVEKQ	172
Qy	169	FF-----INGSNWEGILGLAYAEIARPDSDLPPFDSL VKQTHV-PNLFSLHLCG	217
		: : : : : : : :	
Db	173	IFGEATKQPGVVFI AAKFDGILGMGPYFIS--VNKVL PVFDNL MKQKLVEKNIFS FYL--	228
Qy	218	AGFPLNQSEVLASVGGSMIIGGIDHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMD	277
		: : : : : : : :	
Db	229	-----NRDPTGQPGGELMLGGTDSRYYHGELSYLNVTRKAYWQVHMDQLEV-GSELTL-	280
Qy	278	CKEYNYDKSIVDSGTTNLR LPKKVFEAAVKS IKAASSTEKFPDGFWLGEQLV-CWQAGTT	336
		: : : : : : : : : :	
Db	281	CK--GGCEAIVDTGTSLLVGPVDEVKELQKAIGAVPLIQ-----GEYMIPCEKVSS-	329
Qy	337	PWNIFPVISLYLMGEVTNQSFRTILPQQYLRPVEDVATSQDDCYKFAIS-----Q	387
		: : : : : : :	
Db	330	----LPIITFKLGQ-----NYELHPEKYILKVSQAGKT-----ICLSGFMGMDIPPP	373
Qy	388	SSTGTVMGAVIMEGFYVVFDRARKRIGFAVSA	419
		: : : : :	
Db	374	SGPLWILGDVFIGCYTTFDREYNRVGFAKAA	405

cathepsin E (EC 3.4.23.34) precursor - guinea pig
 N;Alternate names: erythrocyte membrane aspartic proteinase; slow-moving proteinase
 C;Species: Cavia porcellus (guinea pig)
 C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 22-Jun-1999
 C;Accession: A43356
 R;Kageyama, T.; Ichinose, M.; Tsukada, S.; Miki, K.; Kurokawa, K.; Koiwai, O.; Tanji, M.; Yakabe, E.; Athauda, S.B.; Takahashi, K.
 J. Biol. Chem. 267, 16450-16459, 1992
 A;Title: Gastric procathepsin E and progastricsin from guinea pig. Purification, molecular cloning of cDNAs, and characterization of enzymatic properties, with special reference to procathepsin E.
 A;Reference number: A43356; MUID:92355614; PMID:1644829
 A;Accession: A43356
 A;Molecule type: mRNA
 A;Residues: 1-391 <KAG>
 A;Cross-references: GB:M88653; NID:g191294; PIDN:AAA37052.1; PID:g191295
 A;Note: sequence extracted from NCBI backbone (NCBIN:110763, NCBIP:110769)
 C;Superfamily: pepsin
 C;Keywords: aspartic proteinase; hydrolase; membrane protein

Query Match 11.1%; Score 295; DB 2; Length 391;
 Best Local Similarity 26.9%; Pred. No. 6.9e-16;
 Matches 98; Conservative 64; Mismatches 130; Indels 72; Gaps 16;

```

Qy      75 YYVEMTVGSPPQTLNILVDTGSSNFAVGA----APHPFLHRYRQRLSSTYRDLRKGVYV 130
      |: ::||||| :: ||||| | : | : |||||:: :
Db      74 YFGTISIGSPPQNFTVIFDTGSSNLWVPSVYCTSPACQTHPVFHPSLSSTYREVGNFSFSI 133

Qy     131 PYTQGWEGELGTDLVSIHPGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARP 190
      | | | :| | ||: | | : :: | | |:: : ::||| | :|
Db     134 QYGTGSLTGIIGADQVSV-EGLTVVGQQFGESVQEPGKTFVH-AEFDGILGLGYPSLAA- 190

Qy     191 DDSLEPFFDSLVKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSLW 250
      : | ||::: | || : : | : | : || | | :|||
Db     191 -GGVTPVFDNMMAQ----NLVALPM----FSVYMSSNPGGSGSELTFGGYDPSHFSGSLN 241

Qy     251 YTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRLPKKVFEEAAVKSIIK 310
      : | : : |::: : :: | | | | :|||:|: : | : :
Db     242 WVPVTKQAYWQIALDGIQVG--DSVMFCSE--GCQAIVDTGTSLITGP-----PGKIKQLQ 293

Qy     311 AASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRI-----TILPQQ 365
      | : : : | : | : | : | : | : | : | : |
Db     294 EALGATYVDEGY----SVQC-----ANLNMMLDVT---FIINGVPYTLNPTA 333

Qy     366 YLRPVEDVATSQDDCYKFAISQSSTG-----TVMGAVIMEGFYVVFDRARKR 412
      | : | | ||| :|| | : || ||| |
Db     334 Y--TLLDFVDGMQVC-----STGFEGLEIQPPAGPLWILGDVFIRQFYAVFDRGNR 383

Qy     413 IGFA 416
      :| |
Db     384 VGLA 387

```

Search completed: January 21, 2004, 09:26:07
 Job time : 47.9809 secs

OM protein - protein search, using sw model

Run on: January 21, 2004, 09:25:15 ; Search time 100.583 Seconds
(without alignments)
1018.511 Million cell updates/sec

Title: US-09-869-414A-4
Perfect score: 2664
Sequence: 1 MAQALPWLLWLMGAGVLPAA.....CLRCLRQQHDDFADDISLLK 501

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
------------	-------	--------------------	----	-------------

1	2664	100.0	501	9	US-09-794-927-4	Sequence 4, Appli
2	2664	100.0	501	9	US-09-795-847-4	Sequence 4, Appli
3	2664	100.0	501	9	US-09-794-743-4	Sequence 4, Appli
4	2664	100.0	501	9	US-09-794-748-4	Sequence 4, Appli
5	2664	100.0	501	9	US-09-794-925-4	Sequence 4, Appli
6	2664	100.0	501	9	US-09-681-442-4	Sequence 4, Appli
7	2664	100.0	501	11	US-09-869-414-4	Sequence 4, Appli
8	2664	100.0	501	11	US-09-548-366-4	Sequence 4, Appli
9	2656	99.7	501	12	US-10-372-473-1	Sequence 1, Appli
10	2656	99.7	501	15	US-10-032-818-4	Sequence 4, Appli
11	2656	99.7	501	15	US-10-214-932-104	Sequence 104, App
12	2650	99.5	501	11	US-09-969-671A-2	Sequence 2, Appli
13	2650	99.5	501	12	US-10-372-730-9	Sequence 9, Appli
14	2650	99.5	501	15	US-10-308-365-2	Sequence 2, Appli
15	2582	96.9	488	9	US-09-796-264-2	Sequence 2, Appli
16	2582	96.9	488	10	US-09-845-226-2	Sequence 2, Appli
17	2582	96.9	488	10	US-09-795-903A-2	Sequence 2, Appli
18	2582	96.9	488	15	US-10-032-818-2	Sequence 2, Appli
19	2582	96.9	503	9	US-09-796-264-3	Sequence 3, Appli
20	2582	96.9	503	10	US-09-845-226-3	Sequence 3, Appli
21	2582	96.9	503	10	US-09-795-903A-3	Sequence 3, Appli
22	2582	96.9	503	15	US-10-032-818-3	Sequence 3, Appli
23	2567	96.4	501	9	US-09-794-927-8	Sequence 8, Appli
24	2567	96.4	501	9	US-09-795-847-8	Sequence 8, Appli
25	2567	96.4	501	9	US-09-794-743-8	Sequence 8, Appli
26	2567	96.4	501	9	US-09-794-748-8	Sequence 8, Appli
27	2567	96.4	501	9	US-09-794-925-8	Sequence 8, Appli
28	2567	96.4	501	9	US-09-681-442-8	Sequence 8, Appli
29	2567	96.4	501	11	US-09-869-414-8	Sequence 8, Appli
30	2567	96.4	501	11	US-09-548-366-8	Sequence 8, Appli
31	2506.5	94.1	476	9	US-09-794-927-6	Sequence 6, Appli
32	2506.5	94.1	476	9	US-09-795-847-6	Sequence 6, Appli
33	2506.5	94.1	476	9	US-09-794-743-6	Sequence 6, Appli
34	2506.5	94.1	476	9	US-09-794-748-6	Sequence 6, Appli
35	2506.5	94.1	476	9	US-09-794-925-6	Sequence 6, Appli
36	2506.5	94.1	476	9	US-09-681-442-6	Sequence 6, Appli
37	2506.5	94.1	476	11	US-09-869-414-6	Sequence 6, Appli
38	2506.5	94.1	476	11	US-09-548-366-6	Sequence 6, Appli
39	2397	90.0	453	9	US-09-794-927-30	Sequence 30, Appl
40	2397	90.0	453	9	US-09-795-847-30	Sequence 30, Appl
41	2397	90.0	453	9	US-09-794-743-30	Sequence 30, Appl
42	2397	90.0	453	9	US-09-794-748-30	Sequence 30, Appl
43	2397	90.0	453	9	US-09-794-925-30	Sequence 30, Appl
44	2397	90.0	453	9	US-09-681-442-30	Sequence 30, Appl
45	2397	90.0	453	11	US-09-869-414-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1

US-09-794-927-4

; Sequence 4, Application US/09794927

; Patent No. US20010016324A1

; GENERAL INFORMATION:

; APPLICANT: Gurney, Mark E.

```

; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR,
AND
; TITLE OF INVENTION: USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 28341/6280FG
; CURRENT APPLICATION NUMBER: US/09/794,927
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/416,901
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-927-4

```

```

Query Match          100.0%; Score 2664; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 2.6e-253;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRRGSF 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRRGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQT LNI LVD TGSSNFAVGAA PHPFLH RYYQRQLSST 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQT LNI LVD TGSSNFAVGAA PHPFLH RYYQRQLSST 120

Qy    121 YRDLRKG VYVPYTQ GKWEGELGTD LVS I PHGPNVTVRANIAA ITESDKFFINGSNWEGIL 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 YRDLRKG VYVPYTQ GKWEGELGTD LVS I PHGPNVTVRANIAA ITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPD DSLEPFFDSL VKQTHV PNLFS LHL CGAGFPLNQSEVLASVGGSMIIGGI 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GLAYAEIARPD DSLEPFFDSL VKQTHV PNLFS LHL CGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWY YEV IIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DHSLYTGSLWYTPIRREWY YEV IIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300

Qy    301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360

```

```

Qy      361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy      481 RCLRCLRQQHDDFADDISLLK 501
        ||||||||||||||||||
Db      481 RCLRCLRQQHDDFADDISLLK 501

```

RESULT 2

US-09-795-847-4

; Sequence 4, Application US/09795847

; Patent No. US20010018208A1

; GENERAL INFORMATION:

; APPLICANT: Gurney, Mark E.

; APPLICANT: Bienkowski, Michael J.

; APPLICANT: Heinrikson, Robert L.

; APPLICANT: Parodi, Luis A.

; APPLICANT: Yan, Riqiang

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR,
AND

; TITLE OF INVENTION: USES

; TITLE OF INVENTION: THEREFOR

; FILE REFERENCE: 28341/6280DE

; CURRENT APPLICATION NUMBER: US/09/795,847

; CURRENT FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 09/416,901

; PRIOR FILING DATE: 1999-10-13

; PRIOR APPLICATION NUMBER: 60/155,493

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: 09/404,133

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: PCT/US99/20881

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: 60/101,594

; PRIOR FILING DATE: 1998-09-24

; NUMBER OF SEQ ID NOS: 73

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 501

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-795-847-4

Query Match 100.0%; Score 2664; DB 9; Length 501;

Best Local Similarity 100.0%; Pred. No. 2.6e-253;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLPLRSGLGGA PLGLRLPRETDEEPEEPGRGGSF 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLPLRSGLGGA PLGLRLPRETDEEPEEPGRGGSF 60

```

```

Qy      61 VEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120

```

```

Db      61 VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
Qy      121 YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
Db      121 YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
Qy      181 GLAYAEIARPD DSLEPF FDSL VKQTHV PNLFS LHL CGAGFPLNQSEVLASVGGSMIIGGI 240
Db      181 GLAYAEIARPD DSLEPF FDSL VKQTHV PNLFS LHL CGAGFPLNQSEVLASVGGSMIIGGI 240
Qy      241 DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRLPKK 300
Db      241 DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRLPKK 300
Qy      301 VFEEAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
Db      301 VFEEAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
Qy      361 ILPQQYL RPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
Db      361 ILPQQYL RPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
Qy      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
Db      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
Qy      481 RCLRCLRQQHDDFADDISLLK 501
Db      481 RCLRCLRQQHDDFADDISLLK 501

```

RESULT 3

US-09-794-743-4

; Sequence 4, Application US/09794743

; Patent No. US20010021391A1

; GENERAL INFORMATION:

; APPLICANT: Gurney, Mark E.

; APPLICANT: Bienkowski, Michael J.

; APPLICANT: Heinrikson, Robert L.

; APPLICANT: Parodi, Luis A.

; APPLICANT: Yan, Riqiang

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR,
AND

; TITLE OF INVENTION: USES

; TITLE OF INVENTION: THEREFOR

; FILE REFERENCE: 28341/6280BC

; CURRENT APPLICATION NUMBER: US/09/794,743

; CURRENT FILING DATE: 2001-02-27

; PRIOR APPLICATION NUMBER: 09/416,901

; PRIOR FILING DATE: 1999-10-13

; PRIOR APPLICATION NUMBER: 60/155,493

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: 09/404,133

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: PCT/US99/20881

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-743-4

Query Match 100.0%; Score 2664; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 2.6e-253;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MAQALPWLLWLMGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRRGSF 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAQALPWLLWLMGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRRGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPPQTLN ILVDTGSSNFAVGAAPHPFLHRYYQRLSST 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPPQTLN ILVDTGSSNFAVGAAPHPFLHRYYQRLSST 120

Qy    121 YRDLRKG VYVPYTQGWEGELGTDLV SIPHGP NVTVRANIAAITESDKFFINGSNWEGIL 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 YRDLRKG VYVPYTQGWEGELGTDLV SIPHGP NVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPD DSLEPFFDSL VKQTHVPNL FSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GLAYAEIARPD DSLEPFFDSL VKQTHVPNL FSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWY YEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DHSLYTGSLWYTPIRREWY YEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300

Qy    301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNI FPFVISLYLMGEVTNQSF RIT 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNI FPFVISLYLMGEVTNQSF RIT 360

Qy    361 ILPQQYLRPVEDVAT SQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ILPQQYLRPVEDVAT SQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501
        ||||||||||||||||
Db    481 RCLRCLRQQHDDFADDISLLK 501
```

RESULT 4
US-09-794-748-4
; Sequence 4, Application US/09794748
; Patent No. US20020037315A1
; GENERAL INFORMATION:
; APPLICANT: Gurney, Mark E.

```

; APPLICANT: Bienkowski, Michael J.
; APPLICANT: Heinrikson, Robert L.
; APPLICANT: Parodi, Luis A.
; APPLICANT: Yan, Riqiang
; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR,
AND
; TITLE OF INVENTION: USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 28341/6280JL
; CURRENT APPLICATION NUMBER: US/09/794,748
; CURRENT FILING DATE: 2001-02-27
; PRIOR APPLICATION NUMBER: 09/416,901
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-748-4

```

```

Query Match          100.0%;  Score 2664;  DB 9;  Length 501;
Best Local Similarity 100.0%;  Pred. No. 2.6e-253;
Matches 501;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGGSF 60
|
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFAVGAAPHPFLHRYYQRLSST 120
|
Db     61 VEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFAVGAAPHPFLHRYYQRLSST 120

Qy    121 YRDLRKGVPYPYTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180
|
Db    121 YRDLRKGVPYPYTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPDSDLPPFDLSLVKQTHVFNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
|
Db    181 GLAYAEIARPDSDLPPFDLSLVKQTHVFNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK 300
|
Db    241 DHSLYTGSWYTPIRREWYEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK 300

Qy    301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
|
Db    301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360

```

Qy 361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
 |||
 Db 361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 |||
 Db 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy 481 RCLRCLRQQHDDFADDISLLK 501
 |||
 Db 481 RCLRCLRQQHDDFADDISLLK 501

RESULT 5

US-09-794-925-4

; Sequence 4, Application US/09794925

; Patent No. US20020064819A1

; GENERAL INFORMATION:

; APPLICANT: Gurney, Mark E.

; APPLICANT: Bienkowski, Michael J.

; APPLICANT: Heinrikson, Robert L.

; APPLICANT: Parodi, Luis A.

; APPLICANT: Yan, Riqiang

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR, AND USES

; TITLE OF INVENTION: THEREFOR

; FILE REFERENCE: 28341/6280HI

; CURRENT APPLICATION NUMBER: US/09/794,925

; CURRENT FILING DATE: 2001-02-27

; PRIOR APPLICATION NUMBER: 09/416,901

; PRIOR FILING DATE: 1999-10-13

; PRIOR APPLICATION NUMBER: 60/155,493

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: 09/404,133

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: PCT/US99/20881

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: 60/101,594

; PRIOR FILING DATE: 1998-09-24

; NUMBER OF SEQ ID NOS: 73

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 501

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-794-925-4

Query Match 100.0%; Score 2664; DB 9; Length 501;

Best Local Similarity 100.0%; Pred. No. 2.6e-253;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAQALPWLLWLGAGVLPAGTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGGSF 60
 |||

Db 1 MAQALPWLLWLGAGVLPAGTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRGGSF 60

Qy 61 VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAHPFLHRYYQRQLSST 120
 |||

; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-681-442-4

Query Match 100.0%; Score 2664; DB 9; Length 501;
Best Local Similarity 100.0%; Pred. No. 2.6e-253;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRRGSF 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRRGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQT LNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQT LNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy    121 YRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 YRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPDDSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GLAYAEIARPDDSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWYYEVI IVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DHSLYTGSLWYTPIRREWYYEVI IVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300

Qy    301 VFEEAAVKSIIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNI FPFVISLYLMGEVTNQSF RIT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VFEEAAVKSIIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNI FPFVISLYLMGEVTNQSF RIT 360

Qy    361 ILPQQYLRPVEDVATSQDDCYKFAIS QSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ILPQQYLRPVEDVATSQDDCYKFAIS QSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNI PQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNI PQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501
          ||||||||||||||||
Db    481 RCLRCLRQQHDDFADDISLLK 501
```

RESULT 7

US-09-869-414-4

; Sequence 4, Application US/09869414

; Publication No. US20030077226A1

; GENERAL INFORMATION:

; APPLICANT: Beinkowski et al.

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR,
AND USES

```
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 28341/6280M
; CURRENT APPLICATION NUMBER: US/09/869,414
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/416,901
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/155,493
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 09/404,133
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: PCT/US99/20881
; PRIOR FILING DATE: 1999-09-23
; PRIOR APPLICATION NUMBER: 60/101,594
; PRIOR FILING DATE: 1998-09-24
; NUMBER OF SEQ ID NOS: 73
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-869-414-4
```

```
Query Match          100.0%; Score 2664; DB 11; Length 501;
Best Local Similarity 100.0%; Pred. No. 2.6e-253;
Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 MAQALPWLLLLWMGAGVLPAGHTQHGI RPLRSLG GAPLGLRLPRETDEEPEEPGR RGSF 60
        |||
Db      1 MAQALPWLLLLWMGAGVLPAGHTQHGI RPLRSLG GAPLGLRLPRETDEEPEEPGR RGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQT LNILVDTGSSNFAVGAA PHPFLH RYYQRQLSST 120
        |||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQT LNILVDTGSSNFAVGAA PHPFLH RYYQRQLSST 120

Qy    121 YRDLRKGVYVPYTQ GKWEGELGTD LVSIPHGP NVTVRANIAA ITESDKFFINGSNW EGIL 180
        |||
Db    121 YRDLRKGVYVPYTQ GKWEGELGTD LVSIPHGP NVTVRANIAA ITESDKFFINGSNW EGIL 180

Qy    181 GLAYAEIARPD DSLEPF FDSL VKQTHV PNLFS LHL CGAGFPLNQ SEVLASV GGSMIIGGI 240
        |||
Db    181 GLAYAEIARPD DSLEPF FDSL VKQTHV PNLFS LHL CGAGFPLNQ SEVLASV GGSMIIGGI 240

Qy    241 DHSLYTGSLWYTP IRREWY YEVII VRVEINGQ DLKMDCKEY NYDKSIV DSGTTNLR LRPKK 300
        |||
Db    241 DHSLYTGSLWYTP IRREWY YEVII VRVEINGQ DLKMDCKEY NYDKSIV DSGTTNLR LRPKK 300

Qy    301 VFEEAVKSIKAAS STEKFPD GFWLGEQLVCWQAGTTPWN IFPVISLYLMGEV TNQSF RIT 360
        |||
Db    301 VFEEAVKSIKAAS STEKFPD GFWLGEQLVCWQAGTTPWN IFPVISLYLMGEV TNQSF RIT 360

Qy    361 ILPQQYLRPVEDVATS QDDCYKFAISQS SGTVMGAVIMEGFYV VFD RARKRIGFAVSAC 420
        |||
Db    361 ILPQQYLRPVEDVATS QDDCYKFAISQS SGTVMGAVIMEGFYV VFD RARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFV TLDMEDCGYNI PQTDESTLMTIAYVMAA ICALEFMLPLCLMVCQW 480
        |||
Db    421 HVHDEFRTAAVEGPFV TLDMEDCGYNI PQTDESTLMTIAYVMAA ICALEFMLPLCLMVCQW 480
```

Qy 481 RCLRCLRQQHDDFADDISLLK 501
 |||||

D_b 481 RCLRCLRQQHDDFADDISLLK 501

RESULT 8

US-09-548-366-4

; Sequence 4, Application US/09548366

; Publication No. US20030104365A1

; GENERAL INFORMATION:

; APPLICANT: Gurney, Mark E.

; APPLICANT: Bienkowski, Michael J.

; APPLICANT: Heinrikson, Robert L.

; APPLICANT: Parodi, Luis A.

; APPLICANT: Yan, Riqiang

; TITLE OF INVENTION: ALZHEIMER'S DISEASE SECRETASE, APP SUBSTRATES THEREFOR,
AND

; TITLE OF INVENTION: USES THEREFOR

; FILE REFERENCE: 28341/6280A

; CURRENT APPLICATION NUMBER: US/09/548,366

; CURRENT FILING DATE: 2000-04-12

; PRIOR APPLICATION NUMBER: 60/155,493

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: 09/404,133

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: PCT/US99/20881

; PRIOR FILING DATE: 1999-09-23

; PRIOR APPLICATION NUMBER: 60/101,594

; PRIOR FILING DATE: 1998-09-24

; NUMBER OF SEQ ID NOS: 65

```
; SOFTWARE: PatentIn Ver. 2.0
```

; SEQ ID NO 4

```
; LENGTH: 501
```

```
; TYPE: PRT
```

; ORGANISM: Homo sapiens

US-09-548-366-4

Query Match 100.0%; Score 2664; DB 11; Length 501;

Best Local Similarity 100.0%; Pred. No. 2.6e-253;

Matches 501; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MAQALPWLLWMGAGVLP AHGTQH GIRLPLRSGLG GAPLGLRLPRETDEEPEEPGRRGSF 60

Db 1 MAQALPWLLWMGAGVLP AHGTQH GIRLPLRSGLG GAPLGLRLPRETDEEPEEPGRRGSF 60

Qy 61 VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120

[illegible]

Db 61 VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy 121 YRDLRKGVYVPYTQGKWEGLGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180

[illegible]

Db 121 YRDLRKG VYVPYTQ GKWEGELGTDLVSIPHGPNVTVRANIAAITESDKFFINGSNWEGIL 18

Qv 181 GLAYAEIARPDDSLEPFFDSL VKOTHVPNLFSLHLCGAGFPLNOSEVLASVGGSMIIGGI 240

Db 181 GLAYAEIARPDDSLEPFFDSL VKOTHVPNLFSLHLCGAGFPLN QSEVLASVGGSMIIGGI 240

Qy 241 DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK 300
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 241 DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK 300

Qy 301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360

Qy 361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy 481 RCLRCLRQQHDDFADDISLLK 501
 ||||||||||||||||||
 Db 481 RCLRCLRQQHDDFADDISLLK 501

RESULT 9

US-10-372-473-1

; Sequence 1, Application US/10372473
 ; Publication No. US20040005691A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Chou, Kuo-Chen
 ; APPLICANT: Howe, W. Jeffery
 ; TITLE OF INVENTION: Modified BACE
 ; FILE REFERENCE: MBHB 01-1766-A
 ; CURRENT APPLICATION NUMBER: US/10/372,473
 ; CURRENT FILING DATE: 2003-02-21
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 1
 ; LENGTH: 501
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; NAME/KEY: MISC_FEATURE
 ; OTHER INFORMATION: Amino acid sequence of human BACE.

US-10-372-473-1

Query Match 99.7%; Score 2656; DB 12; Length 501;
 Best Local Similarity 99.8%; Pred. No. 1.6e-252;
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MAQALPWLLLWMGAGVLPAGHTQHGI RLPLRSGLGAPLGLRLPRETDEEPEEPGRGGSF 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 MAQALPWLLLWMGAGVLPAGHTQHGI RLPLRSGLGAPLGLRLPRETDEEPEEPGRGGSF 60

Qy 61 VEMVDNLRGKSGQGYVEMTVGSPQTNLN ILVDTGSSNFAVGAAPHPFLHRY YQRQLSST 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 VEMVDNLRGKSGQGYVEMTVGSPQTNLN ILVDTGSSNFAVGAAPHPFLHRY YQRQLSST 120

Qy 121 YRDLRKG VYVPYTQ GKWEGELGTDLV SIPHGPNVTVRANIAAITESDKFFINGSNWEGIL 180

```

      |||
Db      121 YRDLRKGVYVPYTQGKWEDELGTDLVSIHPGNVTVRANIAAATESDKFFINGSNWEGIL 180
Qy      181 GLAYAEIARPDSDLFFDSLKVQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
      |||
Db      181 GLAYAEIARPDSDLFFDSLKVQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
Qy      241 DHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPKK 300
      |||
Db      241 DHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPKK 300
Qy      301 VFEEAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
      |||
Db      301 VFEEAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
Qy      361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
      |||
Db      361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
Qy      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
      |||
Db      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
Qy      481 RCLRCLRQQHDDFADDISLLK 501
      |||
Db      481 RCLRCLRQQHDDFADDISLLK 501

```

RESULT 10

US-10-032-818-4

```

; Sequence 4, Application US/10032818
; Publication No. US20030092629A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Jordan J.N.
; APPLICANT: Koelsch, Gerald
; APPLICANT: Ghosh, Arun K.
; TITLE OF INVENTION: Inhibitors of Memapsin 2 and Use Thereof
; FILE REFERENCE: 2932.1006-007
; CURRENT APPLICATION NUMBER: US/10/032,818
; CURRENT FILING DATE: 2001-12-28
; PRIOR APPLICATION NUMBER: US 60/275,756
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: US 60/258,705
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapien

```

US-10-032-818-4

```

Query Match          99.7%; Score 2656; DB 15; Length 501;
Best Local Similarity 99.8%; Pred. No. 1.6e-252;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy 1 MAQALPWLLLWMGAGVLPAGHTQHGIPLRLSGLGGAPLGLRLPRETDEEPEEPGRRGSF 60

```

      |||
Db      1 MAQALPWLLWLMGAGVLPAGHTQHGI RLP LRSGLGGAPLGLRLPRETDEEPEEPGRG SF 60
Qy      61 VEMVDNLRGKSGQGYVEMTVGSPQTLN ILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
      |||
Db      61 VEMVDNLRGKSGQGYVEMTVGSPQTLN ILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
Qy      121 YRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
      |||
Db      121 YRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
Qy      181 GLAYAEIARPD DSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
      |||
Db      181 GLAYAEIARPD DSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
Qy      241 DHSLYTGSLWYTPIRREWYVEVII VRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
      |||
Db      241 DHSLYTGSLWYTPIRREWYVEVII VRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
Qy      301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360
      |||
Db      301 VFEEAVKSIKAASSTEKFPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360
Qy      361 ILPQQYLRPVEDVATSQDDCYKFAISQS SGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
      |||
Db      361 ILPQQYLRPVEDVATSQDDCYKFAISQS SGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
Qy      421 HVHDEFRTAAVEGPFVTLDMEDCGYNI PQDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
      |||
Db      421 HVHDEFRTAAVEGPFVTLDMEDCGYNI PQDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
Qy      481 RCLRCLRQQHDDFADDISLLK 501
      |||
Db      481 RCLRCLRQQHDDFADDISLLK 501

```

RESULT 11

US-10-214-932-104

; Sequence 104, Application US/10214932

; Publication No. US20030100707A1

; GENERAL INFORMATION:

; APPLICANT: HWANG, Inhwan

; APPLICANT: KIM, Dae Heon

; APPLICANT: LEE, Yong Jik

; TITLE OF INVENTION: SYSTEM FOR DETECTING PROTEASE

; FILE REFERENCE: APB02/US

; CURRENT APPLICATION NUMBER: US/10/214,932

; CURRENT FILING DATE: 2002-08-08

; NUMBER OF SEQ ID NOS: 133

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 104

; LENGTH: 501

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-214-932-104

Query Match

99.7%; Score 2656; DB 15; Length 501;

Best Local Similarity 99.8%; Pred. No. 1.6e-252;
Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRRGSF 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRRGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHFLHRYYQRQLSST 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHFLHRYYQRQLSST 120

Qy    121 YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPDDSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GLAYAEIARPDDSLEPFFDSL VKQTHVPNLFSLQLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 DHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300

Qy    301 VFEEAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 VFEEAVKSIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360

Qy    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNI PQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNI PQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501
        ||||||||||||||||||
Db    481 RCLRCLRQQHDDFADDISLLK 501
```

RESULT 12

US-09-969-671A-2

```
; Sequence 2, Application US/09969671A
; Publication No. US20030036112A1
; GENERAL INFORMATION:
; APPLICANT: CHAPMAN, CONRAD G.
; APPLICANT: MURPHY, KAY
; APPLICANT: POWELL, DAVID J.
; APPLICANT: SMITH, TRUDI S.
; TITLE OF INVENTION: ASP2
; FILE REFERENCE: GH-70368-D1
; CURRENT APPLICATION NUMBER: US/09/969,671A
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: UK 9701684.4
; PRIOR FILING DATE: 1997-01-28
; PRIOR APPLICATION NUMBER: 09/009,191
; PRIOR FILING DATE: 1998-01-20
```

; PRIOR APPLICATION NUMBER: 09/694,200
; PRIOR FILING DATE: 2000-10-23
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-969-671A-2

Query Match 99.5%; Score 2650; DB 11; Length 501;
Best Local Similarity 99.6%; Pred. No. 6.3e-252;
Matches 499; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 MAQALPWL...LWMGAGVLP...AHGTQH...GIRLPLR...SGLGGAPLGLRLPRETDEEPEEPGRRGSF 60
          |||
Db      1 MAQALPWL...LWMGAGVLP...AHGTQH...GIRLPLR...SGLGGAPLGLRLPRETDEEPEEPGRRGSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPPQTLN...ILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120
          |||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPPQTLN...ILVDTGSSNFAVGAAPHPFLHRYYQRQLSST 120

Qy    121 YRDLRKG...VYPYTQ...GKWE...GELGTD...LVSI...PHGPNVTVRANIAA...ITESDKFFINGSNWEGIL 180
          |||
Db    121 YRDLRKG...VYEPYTQ...GKWE...GELGTD...LVSI...PHGPNVTVRANIAA...ITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIAR...PDDSL...EPFFDSL...VKQTHVP...NLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
          |||
Db    181 GLAYAEIAR...PDDSL...EPFFDSL...VKQTHVP...NLFSLQLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWY...EVIIVR...VEINGQDLKMDCKEYNYDKSIVDSGTTNLR...LPKK 300
          |||
Db    241 DHSLYTGSLWYTPIRREWY...EVIIVR...VEINGQDLKMDCKEYNYDKSIVDSGTTNLR...LPKK 300

Qy    301 VFEEAVKSIKAASSTEKFPDGF...LWGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF...RIT 360
          |||
Db    301 VFEEAVKSIKAASSTEKFPDGF...LWGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF...RIT 360

Qy    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
          |||
Db    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNI...PQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
          |||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNI...PQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501
          |||
Db    481 RCLRCLRQQHDDFADDISLLK 501
```

RESULT 13

US-10-372-730-9

; Sequence 9, Application US/10372730
; Publication No. US20030167486A1
; GENERAL INFORMATION:
; APPLICANT: Jacobsen, Helmut

```
; APPLICANT: Mosbach-Ozmen, Laurence
; APPLICANT: Nellboeck-Hochstetter, Peter
; TITLE OF INVENTION: Double transgenic animal model for Alzheimer's Disease
; FILE REFERENCE: Case 21132
; CURRENT APPLICATION NUMBER: US/10/372,730
; CURRENT FILING DATE: 2003-02-24
; PRIOR APPLICATION NUMBER: EP02004331.1
; PRIOR FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-372-730-9
```

```
Query Match          99.5%; Score 2650; DB 12; Length 501;
Best Local Similarity 99.6%; Pred. No. 6.3e-252;
Matches 499; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGLGAPLGLRLPRETDEEPEEPGRRSF 60
      |||
Db      1 MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGLGAPLGLRLPRETDEEPEEPGRRSF 60

Qy     61 VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAPHPFLHRYYQRLSST 120
      |||
Db     61 VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAPHPFLHRYYQRLSST 120

Qy    121 YRDLRKGVPYPTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180
      |||
Db    121 YRDLRKGVPYPTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy    181 GLAYAEIARPDSDLPPFFDSLKVQTHVPLNLSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
      |||
Db    181 GLAYAEIARPDSDLPPFFDSLKVQTHVPLNLSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240

Qy    241 DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPK 300
      |||
Db    241 DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPK 300

Qy    301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRI 360
      |||
Db    301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRI 360

Qy    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
      |||
Db    361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPTDESTLMTIAYVMAAICALEMLPLCLMVCQW 480
      |||
Db    421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPTDESTLMTIAYVMAAICALEMLPLCLMVCQW 480

Qy    481 RCLRCLRQQHDDFADDISLLK 501
      |||
Db    481 RCLRCLRQQHDDFADDISLLK 501
```

RESULT 14
 US-10-308-365-2
 ; Sequence 2, Application US/10308365
 ; Publication No. US20030109022A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CHAPMAN, CONRAD G.
 ; APPLICANT: MURPHY, KAY
 ; APPLICANT: POWELL, DAVID J.
 ; APPLICANT: SMITH, TRUDI S.
 ; TITLE OF INVENTION: ASP 2
 ; FILE REFERENCE: GH-70368-2
 ; CURRENT APPLICATION NUMBER: US/10/308,365
 ; CURRENT FILING DATE: 2002-12-03
 ; PRIOR APPLICATION NUMBER: US/09/694,200
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: UK 9701684.4
 ; PRIOR FILING DATE: 1997-01-28
 ; PRIOR APPLICATION NUMBER: 09/009,191
 ; PRIOR FILING DATE: 1998-01-20
 ; NUMBER OF SEQ ID NOS: 6
 ; SOFTWARE: FastSEQ for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 501
 ; TYPE: PRT
 ; ORGANISM: HOMO SAPIENS
 US-10-308-365-2

Query Match 99.5%; Score 2650; DB 15; Length 501;
 Best Local Similarity 99.6%; Pred. No. 6.3e-252;
 Matches 499; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRRGSF	60
Db	1	MAQALPWLLLWMGAGVLPAGHTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRRGSF	60
Qy	61	VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST	120
Db	61	VEMVDNLRGKSGQGYVEMTVGSPQTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSST	120
Qy	121	YRDLRKGVPYPTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL	180
Db	121	YRDLRKGVEPYPTQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGIL	180
Qy	181	GLAYAEIARPDSDLFFDSLQTHVFNLSLHLCGAGFPLNQSEVLASVGGSMIIGGI	240
Db	181	GLAYAEIARPDSDLFFDSLQTHVFNLSLQLCGAGFPLNQSEVLASVGGSMIIGGI	240
Qy	241	DHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK	300
Db	241	DHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKK	300
Qy	301	VFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT	360
Db	301	VFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT	360
Qy	361	ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC	420

Db 361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 |||

Db 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy 481 RCLRCLRQQHDDFADDISLLK 501
 |||

Db 481 RCLRCLRQQHDDFADDISLLK 501

RESULT 15

US-09-796-264-2

; Sequence 2, Application US/09796264

; Patent No. US20020049303A1

; GENERAL INFORMATION:

; APPLICANT: Tang, Jordan J.N.

; APPLICANT: Lin, Xinli

; APPLICANT: Koelsch, Gerald

; TITLE OF INVENTION: Catalytically Active Recombinant Memapsin and Methods

; TITLE OF INVENTION: of Use Thereof

; FILE REFERENCE: OMRF 179

; CURRENT APPLICATION NUMBER: US/09/796,264

; CURRENT FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: 09/604,608

; PRIOR FILING DATE: 2000-06-27

; PRIOR APPLICATION NUMBER: 60/168,060

; PRIOR FILING DATE: 1999-11-30

; PRIOR APPLICATION NUMBER: 60/177,836

; PRIOR FILING DATE: 2000-01-25

; PRIOR APPLICATION NUMBER: 60/178,368

; PRIOR FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: 60/210,292

; PRIOR FILING DATE: 2000-06-08

; NUMBER OF SEQ ID NOS: 31

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 488

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Purified Memapsin 2

; OTHER INFORMATION: Amino Acids 28-48 are remnant putative propeptide

; OTHER INFORMATION: residues

; OTHER INFORMATION: Amino Acids 58-61, 78, 80, 82-83, 116, 118-121,

; OTHER INFORMATION: 156, 166, 174, 246, 274, 276, 278-281, 283, and

; OTHER INFORMATION: 376-377 are residues in contact with the OM99-2

; OTHER INFORMATION: inhibitor

; OTHER INFORMATION: Amino acids 54-57, 61-68, 73-80, 86-89, 109-111,

; OTHER INFORMATION: 113-118, 123-134, 143-154, 165-168, 198-202, and

; OTHER INFORMATION: 220-224 are N-lobe Beta Strands

; OTHER INFORMATION: Amino Acids 184-191 and 210-217 are N-lobe Helices

; OTHER INFORMATION: Amino acids 237-240, 247-249, 251-256, 259-260,

; OTHER INFORMATION: 273-275, 282-285, 316-318, 331-336, 342-348,

; OTHER INFORMATION: 354-357, 366-370, 372-375, 380-383, 390-395,

; OTHER INFORMATION: 400-405, and 418-420 are C-lobe Beta Strands

; OTHER INFORMATION: Amino Acids 286-299, 307-310, 350-353, 384-387,

; OTHER INFORMATION: and 427-431 are C-lobe Helices
US-09-796-264-2

Query Match 96.9%; Score 2582; DB 9; Length 488;
Best Local Similarity 99.8%; Pred. No. 3e-245;
Matches 487; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	14	AGVLP AHGTQH GIRLPLR SGLGGAPLGLRLPRETDEEPEEPGRRGSFVEMVDNLRGKSGQ	73
Db	1	AGVLP AHGTQH GIRLPLR SGLGGAPLGLRLPRETDEEPEEPGRRGSFVEMVDNLRGKSGQ	60
Qy	74	GYV VEMTVGSPPQTLN ILVDTGSSNFAVGAAPHPFLHRYYQRQLSSTYRDLRKGVVVPYT	133
Db	61	GYV VEMTVGSPPQTLN ILVDTGSSNFAVGAAPHPFLHRYYQRQLSSTYRDLRKGVVVPYT	120
Qy	134	QGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDDS	193
Db	121	QGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDDS	180
Qy	194	LEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSLWYTP	253
Db	181	LEPFFDSL VKQTHVPNLFSLQLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSLWYTP	240
Qy	254	IRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKKVFEAAVKS IKAAS	313
Db	241	IRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKKVFEAAVKS IKAAS	300
Qy	314	STEKFDPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRTILPQQYLRPVEDV	373
Db	301	STEKFDPD GFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRTILPQQYLRPVEDV	360
Qy	374	ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSACHVHDEFRTAAVEG	433
Db	361	ATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSACHVHDEFRTAAVEG	420
Qy	434	PFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQWRCLRCLRQQHDDF	493
Db	421	PFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQWRCLRCLRQQHDDF	480
Qy	494	ADDISLLK	501
Db	481	ADDISLLK	488

Search completed: January 21, 2004, 09:41:41
Job time : 101.583 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 09:16:19 ; Search time 103.457 Seconds
(without alignments)
1249.644 Million cell updates/sec

Title: US-09-869-414A-4
Perfect score: 2664
Sequence: 1 MAQALPWLLLWMGAGVLPAAH.....CLRCLRQQHDDFADDISLLK 501

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query					
No.	Score	Match	Length	DB	ID	Description

1	2650	99.5	501	4	Q8IYC8	Q8iyc8 homo sapien
2	2566	96.3	501	11	Q8C7R1	Q8c7r1 mus musculu
3	2562	96.2	501	11	Q8BQY4	Q8bqy4 mus musculu
4	2478.5	93.0	532	4	Q9ULS1	Q9uls1 homo sapien
5	2374	89.1	467	11	Q8C4F4	Q8c4f4 mus musculu
6	1412	53.0	267	11	Q9CUU5	Q9cuu5 mus musculu
7	1156	43.4	514	11	Q8C5E9	Q8c5e9 mus musculu
8	1155.5	43.4	439	4	Q9H2V8	Q9h2v8 homo sapien
9	1150	43.2	514	11	Q9JL18	Q9jl18 mus musculu
10	1150	43.2	514	11	Q8C793	Q8c793 mus musculu
11	1072.5	40.3	423	4	Q8N2D4	Q8n2d4 homo sapien
12	974.5	36.6	468	4	Q9NZL2	Q9nzl2 homo sapien
13	969.5	36.4	396	4	Q9NZL1	Q9nzl1 homo sapien
14	712.5	26.7	213	4	Q9P0D2	Q9p0d2 homo sapien
15	596.5	22.4	255	11	Q9R1P7	Q9rlp7 mus musculu
16	354.5	13.3	244	5	Q8WQY9	Q8wqy9 aphrocallis
17	345	13.0	76	4	Q8N698	Q8n698 homo sapien
18	335.5	12.6	391	5	Q9VKP6	Q9vvp6 drosophila
19	335	12.6	354	5	Q9GYX7	Q9gyx7 boophilus m
20	319	12.0	384	13	Q9DEC2	Q9dec2 xenopus lae
21	313.5	11.8	385	13	Q9DEC4	Q9dec4 rana catesb
22	312.5	11.7	386	6	Q9BGU5	Q9bgu5 bos taurus
23	311	11.7	387	6	Q9GMY8	Q9gmy8 sorex ungui
24	310	11.6	372	5	Q9VLK3	Q9vlk3 drosophila
25	308	11.6	386	6	Q9GMY7	Q9gmy7 rhinolophus
26	307.5	11.5	383	13	Q9DEC3	Q9dec3 xenopus lae
27	307.5	11.5	387	13	Q9DDV5	Q9ddv5 salvelinus
28	307	11.5	387	6	Q9GMY9	Q9gmy9 suncus muri
29	306.5	11.5	383	13	Q9DE45	Q9de45 salvelinus
30	305.5	11.5	376	13	Q9PUR8	Q9pur8 pseudopleur
31	305	11.4	384	13	Q91322	Q91322 rana catesb
32	304	11.4	382	13	Q9PRG9	Q9prg9 gallus gall
33	304	11.4	423	5	Q9VKP7	Q9vvp7 drosophila
34	298.5	11.2	386	6	Q9GMY6	Q9gmy6 canis famil
35	296.5	11.1	396	13	Q93428	Q93428 chionodraco
36	295.5	11.1	398	13	Q8JH28	Q8jh28 brachydanio
37	295.5	11.1	398	13	Q8AWD9	Q8awd9 brachydanio
38	293.5	11.0	381	6	Q9GK11	Q9gk11 camelus dro
39	293	11.0	399	13	Q93458	Q93458 podarcis si
40	290.5	10.9	380	6	Q28950	Q28950 sus scrofa
41	289.5	10.9	399	13	Q9DD89	Q9dd89 brachydanio
42	288	10.8	387	6	Q8MJU4	Q8mju4 oryctolagus
43	287.5	10.8	444	5	Q21966	Q21966 caenorhabdi
44	287	10.8	427	5	P91802	P91802 schistosoma
45	286.5	10.8	378	13	Q9PUR9	Q9pur9 pseudopleur

ALIGNMENTS

RESULT 1

Q8IYC8

ID Q8IYC8 PRELIMINARY; PRT; 501 AA.

AC Q8IYC8;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Beta-site APP-cleaving enzyme.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; BC036084; AAH36084.1; -.
 SQ SEQUENCE 501 AA; 55824 MW; 768595CF5517EFB7 CRC64;

Query Match 99.5%; Score 2650; DB 4; Length 501;
 Best Local Similarity 99.6%; Pred. No. 1.7e-210;
 Matches 499; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	MAQALPWLLLWMGAGVLPAGHTQHGI	RLPLRSGLG	APLGLRLPRETDEEPEEPGR	RRGSF	60
Db	1	MAQALPWLLLWMGAGVLPAGHTQHGI	RLPLRSGLG	APLGLRLPRETDEEPEEPGR	RRGSF	60
Qy	61	VEMVDNLRGKSGQGYVEMTVGSP	QTLNILVD	TGSSNFAVGAAPHPFLHRY	YQRLSST	120
Db	61	VEMVDNLRGKSGQGYVEMTVGSP	QTLNILVD	TGSSNFAVGAAPHPFLHRY	YQQLFST	120
Qy	121	YRDLRKGVPYPYTQGWEGELG	TDLVSI	PHGPNVTVRANIAAITE	SDKFFINGSN	180
Db	121	YRDLRKGVPYPYTQGWEGELG	TDLVSI	PHGPNVTVRANIAAITE	SDKFFINGSN	180
Qy	181	GLAYAEIARPD	DSLEPF	FD	SLVKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI	240
Db	181	GLAYAEIARPD	DSLEPF	FD	SLVKQTHVPNLFSLQLCGAGFPLNQSEVLASVGGSMIIGGI	240
Qy	241	DHSLYTGSLWYTPIRREWY	YEVII	VRVEINGQDLKMDCKEY	NYDKSIVDSGTTNLR	300
Db	241	DHSLYTGSLWYTPIRREWY	YEVII	VRVEINGQDLKMDCKEY	NYDKSIVDSGTTNLR	300
Qy	301	VFEAAVKSIIKAASSTEK	FPDGF	WLGEQLVCWQAGTTPWNI	FPVISLYLMGEVTNQ	360
Db	301	VFEAAVKSIIKAASSTEK	FPDGF	WLGEQLVCWQAGTTPWNI	FPVISLYLMGEVTNQ	360
Qy	361	ILPQQYLRPVEDVATSQDDCY	KFAISQ	SSTGTVMGAVIMEGFYV	VFDRARKRIGFAVSAC	420
Db	361	ILPQQYLRPVEDVATSQDDCY	KFAISQ	SSTGTVMGAVIMEGFYV	VFDRARKRIGFAVSAC	420
Qy	421	HVHDEFRTAAVEGPFV	TLDMEDCGYNIPQ	TDESTLMTIAYVMAAIC	ALFMLPLCLMVCQW	480
Db	421	HVHDEFRTAAVEGPFV	TLDMEDCGYNIPQ	TDESTLMTIAYVMAAIC	ALFMLPLCLMVCQW	480
Qy	481	RCLRCLRQQHDDFADDIS	LLK			501
Db	481	RCLRCLRQQHDDFADDIS	LLK			501

RESULT 2

Q8C7R1

ID Q8C7R1 PRELIMINARY; PRT; 501 AA.

Qy 481 RCLRCLRQQHDDFADDISLLK 501
||||| |||||||

Dp 481 RCLRCLRHOHDDFADDISLLK 501

RESULT 3

Q8BQY4

```

ID      Q8BQY4          PRELIMINARY;          PRT;      501 AA.
AC      Q8BQY4;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Beta-site APP cleaving enzyme.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/6J; TISSUE=Brain;
RX      MEDLINE=22354683; PubMed=12466851;
RA      The FANTOM Consortium,
RA      the RIKEN Genome Exploration Research Group Phase I & II Team;
RT      "Analysis of the mouse transcriptome based on functional annotation of
RT      60,770 full-length cDNAs.";
RL      Nature 420:563-573(2002).
DR      EMBL; AK046175; BAC32620.1; -.
SQ      SEQUENCE      501 AA;  55816 MW;  C0855513145E024E CRC64;

```

Query Match 96.2%; Score 2562; DB 11; Length 501;
Best Local Similarity 96.0%; Pred. No. 3.3e-203;
Matches 481; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy	1	MAQALPWLLLLWMGAGVLP	AHGTQHGIRLPLRSLGGAPLGLRLP	RETDEEPEEPGR	RSF	60	
		: :					
Db	1	MAPALHWLLLLWVGSGMLP	AQGTHLGI RLPLRSLGAGPPLGLRLP	RETDEESEEPGR	RSF	60	
Qy	61	VEMVDNLRGKSGQGYV	VEMTVGSPQTLN	ILVDTGSSNFAVGA	APHPFLHRY	QRLSST 120	
Db	61	VEMVDNLRGKSGQGYV	VEMTVGSPQTLN	ILVDTGSSNFAVGA	APHPFLHRY	QRLSST 120	
Qy	121	YRDLRKGVVYPYTQ	GKWEGLGTDLVS	IPHGPNVTVRAN	IAAITESDKFF	FINGSNWEGIL 180	
Db	121	YRDLRKGVVYPYTQ	GKWEGLGTDLVS	IPHGPNVTVRAN	IAAITESDKFF	FINGSNWEGIL 180	
Qy	181	GLAYAEIARPD	DSLEPFFDSL	VKQTHVPNLFSL	HLGAGFPLNQ	SEVLASVGGSMIIGGI 240	
			: :	:			
Db	181	GLAYAEIARPD	DSLEPFFDSL	VKQTHIPNIFSL	QLCGAGFPLNQ	TEALASVGGSMIIGGI 240	
Qy	241	DHSLYTGSLWYTP	IRREWYEVII	VRVEINGQDL	KMDCKEYNYDKS	IVDSGTTNLR LPKK 300	
Db	241	DHSLYTGR LWYTP	IRREWYEVII	VRVEINGQDL	KMDCKEYNYDKS	IVDSGTTNLR LPKK 300	
Qy	301	VFEAAVKS	IKAASTEKFPD	GFWLGEQLVCWQ	AGTTPWNIFP	VISLYLMGEVTNQSF	FRIT 360
Db	301	VFEAAVKS	IKAASTEKFPD	GFWLGEQLVCWQ	AGTTPWNIFP	VISLYLMGEVTNQSF	FRIT 360

```

Qy      361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
          |||:|||||
Db      361 ILPQQYLRPVEDVATSQDDCYKFAVSQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
          ||| |||
Db      421 HVHDEFRTAAVEGPFVTADMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy      481 RCLRCLRQHHDDFADDISLLK 501
          ||| |||
Db      481 RCLRCLRHHDDFADDISLLK 501

```

RESULT 4

Q9ULS1

```

ID   Q9ULS1          PRELIMINARY;          PRT;          532 AA.
AC   Q9ULS1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein KIAA1149 (Fragment).
GN   KIAA1149.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20039618; PubMed=10574461;
RA   Hirosawa M., Nagase T., Ishikawa K., Kikuno R., Nomura N., Ohara O.;
RT   "Characterization of cDNA clones selected by the GeneMark analysis
RT   from size-fractionated cDNA libraries from human brain.";
RL   DNA Res. 6:329-336(1999).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
DR   EMBL; AB032975; BAA86463.2; -.
DR   HSSP; P56272; 1AM5.
DR   InterPro; IPR001461; AspproteaseA1.
DR   InterPro; IPR001969; Aspprotease_site.
DR   Pfam; PF00026; asp; 1.
DR   PRINTS; PR00792; PEPSIN.
DR   PROSITE; PS00141; ASP_PROTEASE; 1.
KW   Hypothetical protein; Aspartyl protease; Hydrolase; Protease.
FT   NON_TER      1      1
SQ   SEQUENCE      532 AA;  58720 MW;  98B135D0D5FBD2E8 CRC64;

```

```

Query Match          93.0%;  Score 2478.5;  DB 4;  Length 532;
Best Local Similarity 96.1%;  Pred. No. 2.9e-196;
Matches 473;  Conservative 1;  Mismatches 15;  Indels 3;  Gaps 2;

```

```

Qy      11 WMGAGVLP-AHGTQHGIRLPLRSLGGAPLGLRLPRETDEEPEEPGRRGSFVEMVDNLRG 69
          |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      43 WARECCLPTAPSTASG--CPCAAAGGAPLGLRLPRETDEEPEEPGRRGSFVEMVDNLRG 100

Qy      70 KSGQGYVEMTVGSPPQTLNVLVDTGSSNFAVGAAPHPFLHRYYQRQLSSTYRDLRKGVY 129
          ||| |||
Db      101 KSGQGYVEMTVGSPPQTLNVLVDTGSSNFAVGAAPHPFLHRYYQRQLSSTYRDLRKGVY 160

```

Qy 130 VPYTQGKWEGLGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIAR 189
 |||
 Db 161 VPYTQGKWEGLGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIAR 220
 |||
 Qy 190 PDDSLEPFFDSLQTHVNLFSHLGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSL 249
 |||
 Db 221 PDDSLEPFFDSLQTHVNLFSLQLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSL 280
 |||
 Qy 250 WYTPIRREWYVEIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKKVFEAAVKSI 309
 |||
 Db 281 WYTPIRREWYVEIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKKVFEAAVKSI 340
 |||
 Qy 310 KAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRTILPQQYLRP 369
 |||
 Db 341 KAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRTILPQQYLRP 400
 |||
 Qy 370 VEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSACHVHDEFRTA 429
 |||
 Db 401 VEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSACHVHDEFRTA 460
 |||
 Qy 430 AVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQWRCLRCLRQQ 489
 |||
 Db 461 AVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQWRCLRCLRQQ 520
 |||
 Qy 490 HDDFADDISLLK 501
 |||
 Db 521 HDDFADDISLLK 532

RESULT 5

Q8C4F4

ID Q8C4F4 PRELIMINARY; PRT; 467 AA.
 AC Q8C4F4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Beta-site APP cleaving enzyme.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL; AK082317; BAC38462.1; -.
 SQ SEQUENCE 467 AA; 52063 MW; 31AB674FF1843652 CRC64;

Query Match 89.1%; Score 2374; DB 11; Length 467;
 Best Local Similarity 89.4%; Pred. No. 1e-187;
 Matches 448; Conservative 7; Mismatches 12; Indels 34; Gaps 1;

RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 RT 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Brain;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 RT sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 DR EMBL; AK014390; BAB29317.2; -.

FT NON_TER 1 1
SQ SEQUENCE 267 AA; 30333 MW; 9413EB4530AB63B0 CRC64;

Query Match 53.0%; Score 1412; DB 11; Length 267;
Best Local Similarity 98.9%; Pred. No. 1.6e-108;
Matches 264; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy      235 MIIGGIDHSLYTGSLWYTPIRREWYVEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTN 294
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MIIGGIDHSLYTGSLWYTPIRREWYVEVIIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTN 60

Qy      295 LRLPKKVFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTN 354
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 LRLPKKVFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTN 120

Qy      355 QSFRITILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIG 414
          ||||||||||||||||||||||||||||:||||||||||||||||||||||||||
Db      121 QSFRITILPQQYLRPVEDVATSQDDCYKFAVSQSSTGTVMGAVIMEGFYVVFDRARKRIG 180

Qy      415 FAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLC 474
          ||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db      181 FAVSACHVHDEFRTAAVEGPFVTADMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLC 240

Qy      475 LMVCQWRCLRCLRQHQHDDFADDISLLK 501
          ||||||||||||| ||||||||||||
Db      241 LMVCQWRCLRCLRHQHDDFADDISLLK 267
```

RESULT 7
Q8C5E9

ID Q8C5E9 PRELIMINARY; PRT; 514 AA.
AC Q8C5E9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Beta-site APP-cleaving enzyme 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK078770; BAC37384.1; -.
SQ SEQUENCE 514 AA; 55811 MW; CBB9237BB68A0B2E CRC64;

Query Match 43.4%; Score 1156; DB 11; Length 514;
Best Local Similarity 47.5%; Pred. No. 6.3e-87;
Matches 235; Conservative 77; Mismatches 149; Indels 34; Gaps 9;

```
Qy      2 AQALPWLLLLWMGAGV-----LP---AHGTQHGI RLPLRSGLG--GAP----LGLRL 43
```

```

      | | | | : : | | | | | | | | | |
Db      7 ALLLPVLAQWLLSAVPALAPAPFTLPLQVARATNH--RASAVPGLGTPGLPRADGLALAL 64

Qy      44 PRETDEEPEEPGR-RGSFVEMVDNLRGKSGQGYVEMTVGSPPQTLNILVDTGSSNFAVG 102
      | | | | : | : | | | | | | | | | | : | : | | | | | | | | | |
Db      65 -----EPVRATANFLAMVDNLQGDSEGRGYYLEMLIGTPPQKVQILVDTGSSNFAVA 115

Qy      103 AAPHPFLHRYYQRQLSSTYRDLRKGVPYPTQGKWEGLGTDLVSIPHGPNVTVRANIAA 162
      | | | : : | : : | | | | | | | | | | : | | | : | | : | | |
Db      116 GAPHSYIDTYFDSESSSTYHSGKGFDTVKYTQGSWTGFVGEDLVTPKGFNSFLVNIAT 175

Qy      163 ITESDKFFINGSNWEGILGLAYAEIARPPDSLEPPFDSLQKTHVFNLSLHLCGAGFPL 222
      | | : | : | | | | | | : | : | | | | | | : | : | : | | | :
Db      176 IFESENFFLPGIKWNGILGLAYAAALAKPSSSLETFFDSLVAQAKIPDIFSMQMCAGLPV 235

Qy      223 NQSEVLASVGSGSMIIGGIDHSLYTGSWYTPIRREWYEVIIVRVEINGQDLKMDCKEYN 282
      | : | | : : | | : | | : | | : | | : | : | | | : | : | |
Db      236 AGS---GTNGGSLVLGGIEPSLYKGDIWYTPIKEEWYQIEILKLEIGGQNLNLDCREYN 292

Qy      283 YDKSIVDSGTTNLRRLPKKVFEAAVKSIIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFP 342
      | | : | | | | | | | | : | : | | : | | | | | | | | | | | |
Db      293 ADKAIVDSGTTLLRLPQKVFDVAVVEAVARTSLIPEFSDGFWTGAQLACWTNSETPWAYFP 352

Qy      343 VISLYLMGEVTNQSFRTILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGF 402
      | | : | | : | | | | | | | : : | : : | | | : | | : | | |
Db      353 KISIIYLRDENASRSFRITILPQLYIQPMMGAGFNY-ECYRFGISSSTNALVIGATVMEGF 411

Qy      403 YVVFDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVM 462
      | | | | | : : | | | | | : : | | | | : : | | : | : | :
Db      412 YVVFDRARQRRVGFVAVSPCAEIEGTTVSEISGPFSTEDIASNCVPAQALNEPILWIVSYAL 471

Qy      463 AAICALFMLPLCLMV 477
      : : | : | | : :
Db      472 MSVCGAILLVLLILL 486

```

RESULT 8

Q9H2V8

```

ID   Q9H2V8          PRELIMINARY;          PRT;          439 AA.
AC   Q9H2V8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   CDA13.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pheochromocytoma;
RA   Li Y., Huang Q., Peng, y, Song H., Yu Y., Xu S., Ren S., Chen Z.,
RA   Han Z.;
RL   Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
DR   EMBL; AF212252; AAG41783.1; -.
DR   HSSP; P00797; 2REN.

```


08C793

Query Match 43.2%; Score 1150; DB 11; Length 514;
Best Local Similarity 47.3%; Pred. No. 2e-86;
Matches 233; Conservative 77; Mismatches 153; Indels 30; Gaps 8;

Qy	2	AQALPWLLLLWMGAGV-----LP---AHGTQHGI R LPLRSGLGGAPL-----GLRLPR	45
Db	7	ALLLPVLAQWLLSAVPALAPAPFTLPLQVARATNH--RASAVPGLGTPELPRADGLALAL	64
Qy	46	ETDEEPEEPGR-RGSFVEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFAVGAA	104
Db	65	-----EPVRATANFLAMVDNLQGDSSRGYYLEMLIGTPPKVQVILVDTGSSNFAVAGA	117
Qy	105	PHPFLHRYYQRQLSSTYRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAIT	164
Db	118	PHSYIDTYFDSSESSSTYHSGKGFVDVTVKYTGQSWTGFVGEDLVTIPKGFNSSFLVNIATIF	177
Qy	165	ESDKFFINGSNWEGILGLAYAEIARPDDSLEPFDSL VKQTHVPNLFSLHLCGAGFPLNQ	224
Db	178	ESENFFLPGIKWNGILGLAYAAALAKPSSSLETFFDSLVAQAKIPDIFSMQMCAGLPVAG	237
Qy	225	SEVLASVGGSMIIGGIDHSlyTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYD	284
Db	238	S---GTNGGSLVLGGIEPSLYKGDIWYTPIKEEWYYQIEILKLEIGGQNLNLD CREYNAD	294
Qy	285	KSIVDSGTTNLR LPPKKVFEEAAVKS IKAASSTEKFPDGFWLGEQLVCWQAGTTPWNI FPVI	344
Db	295	KAIVDSGTTLLRLPQKVFD AVVEAVARTSLIPEFS DGFWTGAQLACWTNSETPWAYFPKI	354
Qy	345	SLYLMGEVTNQSFRTITLPQQYL RPVEDVATSQDDCYKFAISQSSTGTMGAVIMEGFYV	404
Db	355	SIYLRDENASRSFR TITLPQLYIQPMMGAGFNY-ECYRFGISSSTNALVIGATVMEGFYV	413
Qy	405	VFDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAA	464

```

          |||||::|:|||| | : : : ||| | |: :| | :| : :
Db      414 VFDRAQRRVGFVSPCAEIEGTTVSEISGPFSTEDIASNCVPAQALNEPILWIVSYALMS 473

Qy      465 ICALFMLPLCLMV 477
          :| :| | |::
Db      474 VCGAILLVLLILL 486

```

RESULT 11

Q8N2D4

```

ID   Q8N2D4          PRELIMINARY;      PRT;    423 AA.
AC   Q8N2D4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein OVARC1000363.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovarian carcinoma;
RA   Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,
RA   Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,
RA   Nagahari K., Sugano S., Isogai T.;
RT   "HRI human cDNA sequencing project.";
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK075539; BAC11682.1; -.
DR   InterPro; IPR001461; AspproteaseA1.
DR   InterPro; IPR001969; Aspprotease_site.
DR   Pfam; PF00026; asp; 2.
DR   PRINTS; PR00792; PEPSIN.
DR   PROSITE; PS00141; ASP_PROTEASE; 2.
KW   Hypothetical protein.
SQ   SEQUENCE 423 AA; 46457 MW; 4D4839F2ED9C2CE1 CRC64;

```

```

Query Match          40.3%; Score 1072.5; DB 4; Length 423;
Best Local Similarity 48.7%; Pred. No. 3.8e-80;
Matches 206; Conservative 74; Mismatches 136; Indels 7; Gaps 4;

```

```

Qy      79 MTVGSPQTLNILVDTGSSNFAVGAAPHPFLHRYYQRQLSSTYRDLRKGVYVPYTQGKWE 138
          | :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1 MLIGTPPQKLQILVDTGSSNFAVAGTPHSYIDTYFDTERSSTYRSKGFDTVKYTGGSWT 60

Qy     139 GELGTDLVSIHPGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDSDLEPFF 198
          | :| | | :| | | : | | | | | | | | | | | | | | | | | | | | | | |
Db      61 GFVGEDLVITIPKGFNTSFLVNIATIFESGNFFLPGIQWNGILGLAYATLAKPSSSLETFF 120

Qy     199 DSLVKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSLSWYTPIRREW 258
          | | | | | :|:|:| | : | | | | : | | | | :|:|:| | | | | | | | | | |
Db     121 DSLVTQANIPNVFSMQMRGAGLPVAGS---GTNGGSLVLGGIEPSLYKGDIIWYTPIKEEW 177

Qy     259 YYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPLPKKVFEAAVKSIIKAASSTEKF 318
          | | :| :|:| | | | :|:|:| | | | | | | | | | | | | | | | | | | |
Db     178 YYQIEILKLEIGGQSLNLDREYNADKAIVDSGTTLLRPLPKKVFDVAVVEAVARASLIPEF 237

```


Db 62 LALE--PALASPAGAAANFLAMVDNLQGDSSGRGYYLEMLIGTPPQKLQILVDTGSSNFAVA 119

Qy 103 AAPHPFLHRYYQRQLSSTYRDLRKGVVYPYTQGWEGELGTDLVSI PHGPNVTVRANIAA 162
 || :: | : : |||| | | |||| | | : | |||:| | | : |||

Db 120 GTPHSYIDTYFDTERSSTYRSKGFDTVTKYTQGSWTGFGEDLVTPKGENTSFLVNIAT 179

Qy 163 ITESDKFFINGSNWEGILGLAYAEIARPPDSLEPFFDSLQVTHVPNLFSLHLCGAGFPL 222
 | ||: ||: | | ||||| :|:| ||| ||||| | :||:| | : |||| | :

Db 180 IFESENFFLPGIKWNGILGLAYATLAKPSSSLETFFDSLVTQANIPNVFSMQMCGAGLPV 239

Qy 223 NQSEVLASVGGSMIIGGIDHSLYTGSLWYTPIRREWYVEIIVRVEINGQDLKMDCKEYN 282
 | : |||::|||: || | :|||: |||: |::| | | :||:| |

Db 240 AGS---GTNGGSLVLGGIEPSLYKGIWYTPIKEEWYQIEILKLEIGGQSLNLD CREYN 296

Qy 283 YDKSIVDSGTTNLRPLPKKVFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFP 342
 ||:||||| |||:| | :|: | |

Db 297 ADKAIVDSGTTLLRLPQKVFDAVVEAVARASLL----- 329

Qy 343 VISLYLMGEVTNQSFRTILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGF 402
 |::| : : ||:| | | : |::| ||||

Db 330 -----YIQPMMGAGLNY-ECYRFGISPSTNALVIGATVMEGF 365

Qy 403 YVVFDRARKRIGFAVSACHVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTIMTIAYVM 462
 ||:||||:| | | | : : || | | : | | :| :

Db 366 YVIFDRAQKRVGFAASPCAIEIAGAAVSEISGPFSTEDVASNCVPAQSLSEPIWIVSYAL 425

Qy 463 AAIC-ALFMLPLCLMVCQWRCLRLRQQHDDFADDISLL 500
 ::| | : : |:: :|| | | : : :| | |

Db 426 MSVCGAILLVLLVLLLLPFRQCR--RPRDPEVVNDESSL 462

RESULT 13

Q9NZL1

ID Q9NZL1 PRELIMINARY; PRT; 396 AA.

AC Q9NZL1;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

DE Aspartyl protease.

GN BACE2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20422477; PubMed=10965118;

RA Solans A., Estivill X., de La Luna S.;

RT "A new aspartyl protease on 21q22.3, BACE2, is highly similar to

RT Alzheimer's amyloid precursor protein beta-secretase.";

RL Cytogenet. Cell Genet. 89:177-184(2000).

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.

DR EMBL; AF188277; AAF35836.1; -.

DR HSSP; P00797; 2REN.

DR InterPro; IPR001461; AspproteaseA1.

DR InterPro; IPR001969; Aspprotease_site.

DR Pfam; PF00026; asp; 1.

DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Aspartyl protease; Hydrolase; Protease.
SQ SEQUENCE 396 AA; 43013 MW; 5023A7AF391CEAC9 CRC64;

Query Match 36.4%; Score 969.5; DB 4; Length 396;
Best Local Similarity 49.3%; Pred. No. 1.1e-71;
Matches 200; Conservative 56; Mismatches 111; Indels 39; Gaps 9;

```
Qy      2 AQALPWLLLWM---GAGVLPAGHTQHGIRLPLRSGLG-----GAPL-----GLR 42
      |  || |  |:      : || |      |||      |  |      ||
Db      7 ALLLPLLAQWLLRAAPELAPAPFT-----LPLRVAAATNRVVAPTGPFGTPAERHADGLA 61

Qy     43 LPRETDEEPEEPGRRGSFVEMVDNLRGKSGQGYVEMTVGSPPQTLNILLVDTGSSNFAVG 102
      |  |      |      :|: |||||:| ||:||||:| | :||| | |||||
Db     62 LALE--PALASPAGAAANFLAMVDNLQGDSEGRGYYLEMLIGTPPQKLQILVDTGSSNFAVA 119

Qy    103 AAPHPFLHRYRQRLSSTYRDLRGVYVPYTQGWEGELGTDLVSIHPGPNVTVRANIAA 162
      || :: | : : |||| | | |||| | | :| |||:| | | : |||
Db    120 GTPHSYIDTYFDTERSSTYRSKGFDTVTKYTQGSWTGFGEDLVITPKGFNTSFLVNIAT 179

Qy    163 ITESDKFFINGSNWEGILGLAYAEIARPDSDLPEFFDSLQTHVFNLFSLHLCGAGFPL 222
      | ||: ||: | | ||||| :|:| ||| ||||| | :||:| | : ||| | :
Db    180 IFESENFFLPGIKWNGILGLAYATLAKPSSSLETFFDSLVTQANIPNVFSMQMCGAGLPV 239

Qy    223 NQSEVLASVGGSMIIGGIDHSLYTGSWYTPIRREWYVEIIVRVEINGQDLKMDCKEYN 282
      |      : |||::||| ||| | :|||: |||:| |::| | | :||:| |
Db    240 AGS---GTNGGSLVLGGIEPSLYKGDWYTPIKEEWYQIEILKLEIGGQSLNLDCREYN 296

Qy    283 YDKSIVDSGTTNLRPLPKKVFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFP 342
      ||:||||| |||:| |:| |:: || :| |||| | || || |||: ||
Db    297 ADKAIVDSGTTLLRPLQKVFDVAVVEAVARASLIPEFSDGFWTGSQACWTNSETPWSYFP 356

Qy    343 VISLYLMGEVTNQSFRTILPQQYLRPVEDVATSQDDCYKF-AISQ 387
      ||:| | | :::||| |||: || :| | | :||
Db    357 KISIYLRDENSRSRFRITILPQK-LRVLQ-----CLKFPGLSQ 393
```

RESULT 14

Q9P0D2

ID Q9P0D2 PRELIMINARY; PRT; 213 AA.
AC Q9P0D2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE HSPC104 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Umbilical cord blood;
RA Zhang Q.H., Ye M., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;
RT "Human partial CDS cloned from cd34+ stem cells."
RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.

DR EMBL; AF161367; AAF28927.1; -.
 DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 FT NON_TER 1 1
 SQ SEQUENCE 213 AA; 24338 MW; EC9D3FA31CFA835C CRC64;

Query Match 26.7%; Score 712.5; DB 4; Length 213;
 Best Local Similarity 83.5%; Pred. No. 7.9e-51;
 Matches 137; Conservative 4; Mismatches 12; Indels 11; Gaps 1;

Qy 238 GGIDHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR 297
 |||||
 Db 1 GGIDHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR 60
 Qy 298 PKKVFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF 357
 |||||
 Db 61 PKKVFEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF 120
 Qy 358 RITILPQQYLRLPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEG 401
 ||||| :| :| || :: |
 Db 121 RITILPQQYLRLP-----WKMWPRPKTTVTVCHLTVIHG 153

RESULT 15

Q9R1P7

ID Q9R1P7 PRELIMINARY; PRT; 255 AA.
 AC Q9R1P7;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Aspartyl protease (Fragment).
 GN BACE2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Accarino M., Fumagalli P., Taramelli R., Ottolenghi S.;
 RT "Cloning of a gene from chromosome 21 Down Region encoding a potential
 RT transmembrane protease."
 RL Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AF051150; AAD45964.1; -.
 DR MEROPS; A01.041; -.
 DR MGD; MGI:1860440; Bace2.
 DR InterPro; IPR001969; Aspprotease_site.
 DR PROSITE; PS00141; ASP_PROTEASE; 1.
 KW Protease.
 FT NON_TER 1 1
 SQ SEQUENCE 255 AA; 28685 MW; 53DE317815996D63 CRC64;

Query Match 22.4%; Score 596.5; DB 11; Length 255;
 Best Local Similarity 47.8%; Pred. No. 4e-41;
 Matches 109; Conservative 44; Mismatches 74; Indels 1; Gaps 1;

Qy 250 WYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLRPKKVFEAAVKS 309
 |||||: |||||: |::|| ||:| :||:| ||:|||| ||||:||||:| |:::

Db 1 WYTPIKEEWYYQIEILKLEIGGQNLNLD CREYNADKAIVDSGTTLLRLPQKVFD AVEAV 60
 Qy 310 KAASSTEKFPDGF LGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RITILPQQYL RP 369
 | :| |||| | || || || || || ||:| | :||| ||||| | :||
 Db 61 ARTSLIPEFSDGF WTGAQLACWTNSETPWAYFPKIS IYLRDENASRSFRITILP QLYIQP 120
 Qy 370 VEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYV VFD RARKRIGFAVSACHVHDEFRTA 429
 : : :||:| || |: ||:| :||| ||||| ||:| :||| | : :
 Db 121 MMGAGFNY-ECYRFGISSSTNALVIGATVMEGFYV VFDRAQRRVGFAVSPCAEIEGTTVS 179
 Qy 430 AVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMV 477
 : ||| | |: :| | :|| : :| :| | | :|
 Db 180 EISGPFSTEDIASNCVPAQALNEPILWIVSYALMSVCGAILLV LILL 227

Search completed: January 21, 2004, 09:25:07
 Job time : 106.457 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 09:15:44 ; Search time 24.9063 Seconds
(without alignments)
945.960 Million cell updates/sec

Title: US-09-869-414A-4
Perfect score: 2664
Sequence: 1 MAQALPWLLLWMGAGVLPAAH.....CLRCLRQQHDDFADDISLLK 501

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	2656	99.7	501	1	BACE_HUMAN	P56817 homo sapien
2	2569	96.4	501	1	BACE_RAT	P56819 rattus norv
3	2567	96.4	501	1	BACE_MOUSE	P56818 mus musculu
4	1173.5	44.1	518	1	BAE2_HUMAN	Q9y5z0 homo sapien
5	330	12.4	324	1	PEP1_GADMO	P56272 gadus morhu
6	314.5	11.8	390	1	CATD_BOVIN	P80209 bos taurus
7	309	11.6	387	1	PEP1_RABIT	P28712 oryctolagus
8	307.5	11.5	388	1	PEP4_MACFU	P27678 macaca fusc
9	305	11.4	367	1	PEPA_CHICK	P00793 gallus gall
10	301.5	11.3	383	1	PEPE_CHICK	P16476 gallus gall
11	301.5	11.3	396	1	CATE_HUMAN	P14091 homo sapien
12	300.5	11.3	412	1	CATD_HUMAN	P07339 homo sapien
13	299	11.2	387	1	PEP2_RABIT	P27821 oryctolagus
14	298	11.2	387	1	PEP4_RABIT	P28713 oryctolagus
15	297	11.1	407	1	CATD_RAT	P24268 rattus norv
16	295	11.1	391	1	CATE_CAVPO	P25796 cavia porce
17	294.5	11.1	388	1	PEP2_MACFU	P27677 macaca fusc

18	289	10.8	387	1	PEP3_RABIT	P27822	oryctolagus
19	288.5	10.8	388	1	PEPA_HUMAN	P00790	homo sapien
20	288.5	10.8	388	1	PEPA_MACMU	P11489	macaca mula
21	288.5	10.8	398	1	CATE_RAT	P16228	rattus norv
22	287	10.8	410	1	CATD_MOUSE	P18242	mus musculu
23	286.5	10.8	388	1	PEP1_MACFU	P03954	macaca fusc
24	286	10.7	398	1	CATD_CHICK	Q05744	gallus gall
25	284.5	10.7	381	1	CHYM_SHEEP	P18276	ovis aries
26	281.5	10.6	386	1	PEPA_PIG	P00791	sus scrofa
27	281	10.5	387	1	PEPA_CALJA	Q9n2d4	callithrix
28	280.5	10.5	396	1	CATD_CLUHA	Q9dex3	clupea hare
29	280.5	10.5	397	1	CATE_MOUSE	P70269	mus musculu
30	276.5	10.4	381	1	CHYM_BOVIN	P00794	bos taurus
31	276.5	10.4	396	1	CATE_RABIT	P43159	oryctolagus
32	274.5	10.3	419	1	CARV_CANAL	P10977	candida alb
33	273.5	10.3	376	1	PAG2_BOVIN	Q28057	bos taurus
34	273.5	10.3	377	1	PEPC_MACFU	P03955	macaca fusc
35	273	10.2	388	1	PEPF_RABIT	P27823	oryctolagus
36	270.5	10.2	381	1	CHYM_CALJA	Q9n2d2	callithrix
37	268	10.1	396	1	CARP_NEUCR	Q01294	neurospora
38	267	10.0	365	1	CATD_SHEEP	Q9mzs8	ovis aries
39	266.5	10.0	388	1	PEPC_CALJA	Q9n2d3	callithrix
40	266	10.0	394	1	PEPC_CAVPO	Q64411	cavia porce
41	266	10.0	405	1	CARP_YEAST	P07267	saccharomyc
42	264.5	9.9	388	1	PEPC_HUMAN	P20142	homo sapien
43	264	9.9	388	1	PAG_HORSE	Q28389	equus cabal
44	262	9.8	496	1	ASPR_ORYSA	P42211	oryza sativ
45	261.5	9.8	387	1	ASPP_AEDAE	Q03168	aedes aegyp

ALIGNMENTS

RESULT 1

BACE_HUMAN

ID BACE_HUMAN STANDARD; PRT; 501 AA.
AC P56817; Q9BYB9; Q9BYC0; Q9BYC1; Q9UJT5;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)
DE (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl
DE protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)
DE (Memapsin-2).
GN BACE OR BACE1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM A).
RC TISSUE=Brain;
RX MEDLINE=20002972; PubMed=10531052;
RA Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
RA Denis P., Teplov D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
RA Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
RA Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,

RA Treanor J., Rogers G., Citron M.;
 RT "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
 RT the transmembrane aspartic protease BACE.";
 RL Science 286:735-741(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM A), SEQUENCE OF 46-68, AND
 RP CHARACTERIZATION.
 RC TISSUE=Brain;
 RX MEDLINE=20057171; PubMed=10591214;
 RA Sinha S., Anderson J.P., Barbour R., Basi G.S., Caccavello R.,
 RA Davis D., Doan M., Dovey H.F., Frigon N., Hong J., Jacobson-Croak K.,
 RA Jewett N., Keim P., Knops J., Lieberburg I., Power M., Tan H.,
 RA Tatsuno G., Tung J., Schenk D., Seubert P., Suomensaaari S.M., Wang S.,
 RA Walker D., Zhao J., McConlogue L., Varghese J.;
 RT "Purification and cloning of amyloid precursor protein beta-secretase
 RT from human brain.";
 RL Nature 402:537-540(1999).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RX MEDLINE=20057170; PubMed=10591213;
 RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
 RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
 RA Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;
 RT "Membrane-anchored aspartyl protease with Alzheimer's disease beta-
 RT secretase activity.";
 RL Nature 402:533-537(1999).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RX MEDLINE=20120043; PubMed=10656250;
 RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,
 RA Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,
 RA Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;
 RT "Identification of a novel aspartic proteinase (Asp 2) as beta-
 RT secretase.";
 RL Mol. Cell. Neurosci. 14:419-427(1999).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RC TISSUE=Brain, and Pancreas;
 RA Michel B., De Pietri Tonelli D., Zacchetti D., Keller P.;
 RT "New beta-site APP cleaving enzyme isoform (BACE-1B) obtained from
 RT human brain and pancreas.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM C).
 RC TISSUE=Pancreas;
 RA Zacchetti D., De Pietri Tonelli D., Schnurbus R.;
 RT "New beta-site APP cleaving enzyme isoform (BACE-1C) obtained from
 RT human pancreas.";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A. (ISOFORMS B; C AND D).
 RC TISSUE=Brain;
 RX MEDLINE=21408467; PubMed=11516562;
 RA Tanahashi H., Tabira T.;
 RT "Three novel alternatively spliced isoforms of the human beta-site
 RT amyloid precursor protein cleaving enzyme (BACE) and their effect on
 RT amyloid beta-peptide production.";

RL Neurosci. Lett. 307:9-12(2001).
 RN [8]
 RP SEQUENCE OF 14-501 FROM N.A. (ISOFORM A), AND CHARACTERIZATION.
 RX MEDLINE=20144060; PubMed=10677483;
 RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;
 RT "Human aspartic protease memapsin 2 cleaves the beta-secretase site of
 RT beta-amyloid precursor protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).
 RN [9]
 RP DISULFIDE BONDS.
 RX MEDLINE=21950860; PubMed=11953458;
 RA Fischer F., Molinari M., Bodendorf U., Paganetti P.;
 RT "The disulphide bonds in the catalytic domain of BACE are critical but
 RT not essential for amyloid precursor protein processing activity.";
 RL J. Neurochem. 80:1079-1088(2002).
 CC -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
 CC AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
 CC THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,
 CC LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
 CC SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
 CC FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=A; Synonyms=BACE-1A, BAC-501;
 CC IsoId=P56817-1; Sequence=Displayed;
 CC Name=B; Synonyms=BACE-1B, BACE-I-476;
 CC IsoId=P56817-2; Sequence=VSP_005223;
 CC Name=C; Synonyms=BACE-1C, BACE-I-457;
 CC IsoId=P56817-3; Sequence=VSP_005222;
 CC Name=D; Synonyms=BACE-1D, BACE-I-432;
 CC IsoId=P56817-4; Sequence=VSP_005222, VSP_005223;
 CC -!- TISSUE SPECIFICITY: BRAIN.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF190725; AAF04142.1; -.
 DR EMBL; AF201468; AAF18982.1; -.
 DR EMBL; AF200343; AAF17079.1; -.
 DR EMBL; AF204943; AAF26367.1; -.
 DR EMBL; AF338816; AAK38374.1; -.
 DR EMBL; AF338817; AAK38375.1; -.
 DR EMBL; AB050436; BAB40931.1; -.
 DR EMBL; AB050437; BAB40932.1; -.
 DR EMBL; AB050438; BAB40933.1; -.
 DR EMBL; AF200193; AAF13715.1; -.
 DR PIR; A59090; A59090.
 DR PDB; 1M4H; 28-AUG-02.
 DR MEROPS; A01.004; -.
 DR Genew; HGNC:933; BACE.

DR MIM; 604252; -.
 DR GO; GO:0005887; C:integral to plasma membrane; TAS.
 DR GO; GO:0008798; F:beta-aspartyl-peptidase activity; TAS.
 DR GO; GO:0009405; P:pathogenesis; TAS.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; TAS.
 DR InterPro; IPR001969; Aspprotease_site.
 DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPSIN.
 DR PROSITE; PS00141; ASP_PROTEASE; 1.
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;
 KW Signal; Alternative splicing; 3D-structure.
 FT SIGNAL 1 21 POTENTIAL.
 FT PROPEP 22 45
 FT CHAIN 46 501 BETA-SECRETASE.
 FT DOMAIN 22 457 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 458 478 POTENTIAL.
 FT DOMAIN 479 501 CYTOPLASMIC (POTENTIAL).
 FT ACT_SITE 93 93 BY SIMILARITY.
 FT ACT_SITE 289 289 BY SIMILARITY.
 FT DISULFID 216 420
 FT DISULFID 278 443
 FT DISULFID 330 380
 FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 146 189 Missing (in isoform C and isoform D).
 FT /FTId=VSP_005222.
 FT VARSPLIC 190 214 Missing (in isoform B and isoform D).
 FT /FTId=VSP_005223.
 SQ SEQUENCE 501 AA; 55763 MW; 377CE4C824ACEF05 CRC64;

Query Match 99.7%; Score 2656; DB 1; Length 501;
 Best Local Similarity 99.8%; Pred. No. 2.7e-206;
 Matches 500; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	MAQALPWLLWLGAGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEEPEEPGRRGSF	60
Db	1	MAQALPWLLWLGAGVLPAGHTQHGIRLPLRSGLGAPLGLRLPRETDEEPEEPGRRGSF	60
Qy	61	VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAPHPFLHRYYQRQLSST	120
Db	61	VEMVDNLRGKSGQGYVEMTVGSPQTNLILVDTGSSNFAVGAAPHPFLHRYYQRQLSST	120
Qy	121	YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWE GIL	180
Db	121	YRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDKFFINGSNWE GIL	180
Qy	181	GLAYAEIARPD DSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI	240
Db	181	GLAYAEIARPD DSLEPFFDSL VKQTHVPNLFSLQLCGAGFPLNQSEVLASVGGSMIIGGI	240
Qy	241	DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK	300
Db	241	DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK	300

```

Qy      301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRLT 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRLT 360

Qy      361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 ILPQQYLRPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy      481 RCLRCLRQQHDDFADDISLLK 501
        ||||||||||||||||||
Db      481 RCLRCLRQQHDDFADDISLLK 501

```

RESULT 2

BACE_RAT

```

ID      BACE_RAT          STANDARD;      PRT;      501 AA.
AC      P56819;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      28-FEB-2003 (Rel. 41, Last annotation update)
DE      Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)
DE      (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl
DE      protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)
DE      (Memapsin-2).
GN      BACE.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxID=10116;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=20002972; PubMed=10531052;
RA      Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
RA      Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
RA      Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
RA      Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,
RA      Treanor J., Rogers G., Citron M.;
RT      "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
RT      the transmembrane aspartic protease BACE.";
RL      Science 286:735-741(1999).
CC      -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
CC      AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF
CC      THE A-BETA PEPTIDE SEQUENCE, BETWEEN RESIDUES 671 AND 672 OF APP,
CC      LEADS TO THE GENERATION AND EXTRACELLULAR RELEASE OF BETA-CLEAVED
CC      SOLUBLE APP, AND A CORRESPONDING CELL-ASSOCIATED CARBOXY-TERMINAL
CC      FRAGMENT WHICH IS LATER RELEASE BY GAMMA-SECRETASE (BY
CC      SIMILARITY).
CC      -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.

```

```

CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its

```



```

Db      301 VFEEAVKSIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSFRT 360
Qy      361 ILPQQYLRLPVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
          |||||:|||||
Db      361 ILPQQYLRLPVEDVATSQDDCYKFAVSQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
Qy      421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
          ||||| |||||
Db      421 HVHDEFRTAAVEGPFVTADMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
Qy      481 RCLRCLRQHHDDFADDISLLK 501
          |||||
Db      481 RCLRCLRHHQHHDDFADDISLLK 501

```

RESULT 3

BACE_MOUSE

```

ID  BACE_MOUSE      STANDARD;      PRT;      501 AA.
AC  P56818;
DT  30-MAY-2000 (Rel. 39, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Beta-secretase precursor (EC 3.4.23.-) (Beta-site APP cleaving enzyme)
DE  (Beta-site amyloid precursor protein cleaving enzyme) (Aspartyl
DE  protease 2) (Asp 2) (ASP2) (Membrane-associated aspartic protease 2)
DE  (Memapsin-2).
GN  BACE.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=20002972; PubMed=10531052;
RA  Vassar R., Bennett B.D., Babu-Khan S., Kahn S., Mendiaz E.A.,
RA  Denis P., Teplow D.B., Ross S., Amarante P., Loeloff R., Luo Y.,
RA  Fisher S., Fuller J., Edenson S., Lile J., Jarosinski M.A.,
RA  Biere A.L., Curran E., Burgess T., Louis J.-C., Collins F.,
RA  Treanor J., Rogers G., Citron M.;
RT  "Beta-secretase cleavage of Alzheimer's amyloid precursor protein by
RT  the transmembrane aspartic protease BACE.";
RL  Science 286:735-741(1999).
RN  [2]
RP  REVISIONS TO 6 AND 81-87.
RA  Bennett B.D., Vassar R., Citron M.;
RL  Submitted (JAN-2000) to the EMBL/GenBank/DDBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=20057170; PubMed=10591213;
RA  Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
RA  Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
RA  Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;
RT  "Membrane-anchored aspartyl protease with Alzheimer's disease
RT  beta-secretase activity.";
RL  Nature 402:533-537(1999).
CC  -!- FUNCTION: RESPONSIBLE FOR THE PROTEOLYTIC PROCESSING OF THE
CC  AMYLOID PRECURSOR PROTEIN (APP). CLEAVES AT THE AMINO TERMINUS OF

```


Qy 121 YRDLRKG VYVPYTQ GKWEGELGTDLV SIPHGPNVTVRANIAAITESDKFFINGSNWEGIL 180
 |||
 Db 121 YRDLRKG VYVPYTQ GKWEGELGTDLV SIPHGPNVTVRANIAAITESDKFFINGSNWEGIL 180

Qy 181 GLAYAEIAR PDDSLEPFFDSL VKQTHVPNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGI 240
 |||:|:| |||
 Db 181 GLAYAEIAR PDDSLEPFFDSL VKQTHIPNIFSLQLCGAGFPLNQTEALASVGGSMIIGGI 240

Qy 241 DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300
 |||
 Db 241 DHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMDCKEYNYDKSIVDSGTTNLR LPKK 300

Qy 301 VFEEAVKSIK AASSTEKFPDGF LGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360
 |||
 Db 301 VFEEAVKSIK AASSTEKFPDGF LGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQSF RIT 360

Qy 361 ILPQQYL R PVEDVATSQDDCYKFAISQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420
 |||:|:| |||
 Db 361 ILPQQYL R PVEDVATSQDDCYKFAVSQSSTGTVMGAVIMEGFYVVFDRARKRIGFAVSAC 420

Qy 421 HVHDEFRTAAVEGPFVTLDMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480
 |||
 Db 421 HVHDEFRTAAVEGPFV TADMEDCGYNIPQTDESTLMTIAYVMAAICALFMLPLCLMVCQW 480

Qy 481 RCLRCLRQHDDFADDISLLK 501
 |||
 Db 481 RCLRCLRHQDDFADDISLLK 501

RESULT 4

BAE2_HUMAN

ID BAE2_HUMAN STANDARD; PRT; 518 AA.
 AC Q9Y5Z0; Q9UJT6;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Beta secretase 2 precursor (EC 3.4.23.-) (Beta-site APP-cleaving
 DE enzyme 2) (Aspartyl protease 1) (Asp 1) (ASPl) (Membrane-associated
 DE aspartic protease 1) (Memapsin-1).
 GN BACE2 OR ASP21.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20057170; PubMed=10591213;
 RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,
 RA Brashier J.R., Stratman N.C., Mathews W.R., Buhl A.E., Carter D.B.,
 RA Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;
 RT "Membrane-anchored aspartyl protease with Alzheimer's disease
 RT beta-secretase activity."
 RL Nature 402:533-537(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RA Xin H., Stephans J.C., Duan X., Harrowe G., Kim E., Grieshammer U.,

RA Giese K.;
 RT "Identification of a novel aspartic-like protease differentially
 RT expressed in human breast cancer cell lines.";
 RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Accarino M.P., Fumagalli P., Ottolenghi S., Taramelli R.;
 RT "Cloning of a gene from chromosome 21 Down region encoding a potential
 RT transmembrane aspartyl protease.";
 RL Submitted (FEB-1998) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Solans A., Estivill X., de la Luna S.;
 RT "Cloning of a novel mammalian aspartyl protease.";
 RL Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20120043; PubMed=10656250;
 RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,
 RA Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,
 RA Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;
 RT "Identification of a novel aspartic proteinase (Asp 2) as
 RT beta-secretase.";
 RL Mol. Cell. Neurosci. 14:419-427(1999).
 RN [6]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20144060; PubMed=10677483;
 RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;
 RT "Human aspartic protease memapsin 2 cleaves the beta-secretase site of
 RT beta-amyloid precursor protein.";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).
 RN [7]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
 RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
 RA Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
 RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
 RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
 RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
 RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
 RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
 RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
 RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
 RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
 RA Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 RN [8]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF200342; AAF17078.1; -.
 DR EMBL; AF117892; AAD45240.1; -.
 DR EMBL; AF050171; AAD45963.1; -.
 DR EMBL; AF178532; AAF29494.1; -.
 DR EMBL; AF204944; AAF26368.1; -.
 DR EMBL; AF200192; AAF13714.1; -.
 DR EMBL; AL163284; CAB90458.1; -.
 DR EMBL; AL163285; CAB90554.1; -.
 DR EMBL; BC014453; AAH14453.1; -.
 DR HSSP; P00797; 2REN.
 DR MEROPS; A01.041; -.
 DR Genew; HGNC:934; BACE2.
 DR MIM; 605668; -.
 DR GO; GO:0005624; C:membrane fraction; TAS.
 DR GO; GO:0006464; P:protein modification; TAS.
 DR GO; GO:0009306; P:protein secretion; TAS.
 DR InterPro; IPR001969; Aspprotease_site.
 DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPSIN.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;
 KW Signal.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 ? POTENTIAL.
 FT CHAIN ? 518 BETA SECRETASE 2.
 FT DOMAIN 21 473 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 474 494 POTENTIAL.
 FT DOMAIN 495 518 CYTOPLASMIC (POTENTIAL).
 FT ACT_SITE 110 110 BY SIMILARITY.
 FT ACT_SITE 303 303 BY SIMILARITY.

OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadidae; Gadus.
 OX NCBI_TaxID=8049;
 RN [1]
 RP SEQUENCE, AND X-RAY CRYSTALLOGRAPHY.
 RC TISSUE=Stomach;
 RA Karlsen S., Hough E., Olsen R.L.;
 RT "Structure and proposed amino-acid sequence of a pepsin from Atlantic
 RT cod (Gadus morhua).";
 RL Acta Crystallogr. D 54:32-46(1998).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 DR PDB; 1AM5; 24-DEC-97.
 DR InterPro; IPR001969; Aspprotease_site.
 DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPSIN.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; 3D-structure.
 FT ACT_SITE 32 32 BY SIMILARITY.
 FT ACT_SITE 214 214 BY SIMILARITY.
 FT DISULFID 45 50 BY SIMILARITY.
 FT DISULFID 206 209 BY SIMILARITY.
 FT DISULFID 247 280 BY SIMILARITY.
 FT STRAND 2 9
 FT TURN 10 12
 FT STRAND 13 20
 FT TURN 21 24
 FT STRAND 25 32
 FT TURN 33 34
 FT STRAND 38 40
 FT STRAND 42 42
 FT TURN 43 44
 FT HELIX 48 51
 FT TURN 52 52
 FT STRAND 56 56
 FT HELIX 58 60
 FT TURN 62 63
 FT STRAND 65 74
 FT STRAND 79 90
 FT STRAND 96 106
 FT TURN 110 114
 FT STRAND 119 122
 FT HELIX 126 128
 FT HELIX 130 132
 FT HELIX 136 142
 FT TURN 143 144
 FT STRAND 150 154
 FT TURN 158 159
 FT STRAND 163 167
 FT HELIX 172 174
 FT STRAND 175 175
 FT STRAND 180 187
 FT TURN 188 189
 FT STRAND 190 194
 FT STRAND 196 199
 FT TURN 200 201
 FT STRAND 202 203
 FT STRAND 209 213

```

FT    TURN          215    216
FT    STRAND        220    222
FT    TURN          224    226
FT    HELIX         227    234
FT    TURN          235    235
FT    STRAND        237    238
FT    STRAND        243    244
FT    TURN          247    248
FT    STRAND        256    260
FT    TURN          261    262
FT    STRAND        263    267
FT    HELIX         269    272
FT    STRAND        273    275
FT    STRAND        280    282
FT    STRAND        284    286
FT    STRAND        296    299
FT    HELIX         301    306
FT    STRAND        307    312
FT    TURN          313    316
FT    STRAND        317    324
SQ    SEQUENCE      324 AA;  34014 MW;  EE3A6097B6941DD7 CRC64;

```

```

Query Match          12.4%; Score 330; DB 1; Length 324;
Best Local Similarity 27.9%; Pred. No. 3.7e-19;
Matches 104; Conservative 67; Mismatches 136; Indels 66; Gaps 15;

```

```

Qy      63 MVDNLRGKSGQGYVEMTVGSPPQTLNILDVTGSSNFAVG----AAPHFPLHRYYQRQLS 118
      : : : : : || : : : : : : : : : : : : : : : : : : : : : :
Db      2 VTEQMKNADTEYYGVISIGTPPESFKVIFDTGSSNLWVSSSHCSAQACSNHNKFKPRQS 61

Qy     119 STYRDLRKGVVYPYTQGWEGELGTDLVSI PHG--PNVTVRANIAAITESDKFFINGSNW 176
      ||| : | | : | | | | | : | | : : : | | : :
Db      62 STYVETGKTVDLTYGTGGMRGILGQD TVSVGGGSDPNQELG---ESQTEPGPFQA-AAPF 117

Qy     177 EGILGLAYAEIARPDDSLEPFFDSL VKQTHV--PNLFSLHLCGAGFPLNQSEVLASVGGSM 235
      : || || || || || || : | : | : | : | : | : | :
Db     118 DGILGLAYPSIAAA--GAVPVFDNMGSQSLVEKD LFSFYLSGGG--ANGSEVM----- 166

Qy     236 IIGGIDHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKMD--CKEYNYDKSIVDSGTTN 294
      : || : | : | || : : | : | : : : : : : : : : : : : : : : :
Db     167 -LGGVDNSHYTGSIHWIPVTAEKYWQVALDGITVNGQTAACEGC-----QAIVDTGTSTK 219

Qy     295 LRLPKKVFEAAVKSIIKAASSTEKFPDGFWLGEQLVCWQAGTTPWNIFPVISLYLMGEVTN 354
      : | : | : | : : : | : | : | : | : | : | : |
Db     220 IVAPVSALANIMKDIGASEN-----QGEMMGN---CASVQSLPDITF-----TI 260

Qy     355 QSFRITILPQQYLRPVEDVATSQDDCYKFAISQSSTGT-----VMGAVIMEGFYVVF 406
      : : | | : : || : : | : | : : | : : : :
Db     261 NGVKQPLPPSAYIEGDQAFCTS-----GLGSSGVPSNTSELWIFGDVFLRNYTTIY 311

Qy     407 DRARKRIGFAVSA 419
      || : : || : :
Db     312 DRTNNKVGAFAPAA 324

```

RESULT 6
 CATD_BOVIN

ID CATD_BOVIN STANDARD; PRT; 390 AA.
 AC P80209; Q9TS27;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cathepsin D precursor (EC 3.4.23.5).
 GN CTSD.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE OF 1-48.
 RC TISSUE=Milk;
 RX MEDLINE=93202276; PubMed=8454061;
 RA Larsen L.B., Boisen A., Petersen T.E.;
 RT "Procathepsin D cannot autoactivate to cathepsin D at acid pH."
 RL FEBS Lett. 319:54-58(1993).
 RN [2]
 RP SEQUENCE OF 45-390, AND X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
 RC TISSUE=Liver;
 RX MEDLINE=93223670; PubMed=8467789;
 RA Metcalf P., Fusek M.;
 RT "Two crystal structures for cathepsin D: the lysosomal targeting
 RT signal and active site."
 RL EMBO J. 12:1293-1302(1993).
 CC -!- FUNCTION: Acid protease active in intracellular protein breakdown.
 CC -!- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,
 CC that of pepsin A. Does not cleave the 4-Gln-|-His-5 bond in B
 CC chain of insulin.
 CC -!- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
 CC -!- SUBCELLULAR LOCATION: Lysosomal.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 DR HSSP; P07339; 1LYB.
 DR MEROPS; A01.009; -.
 DR InterPro; IPR001969; Aspprotease_site.
 DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPSIN.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Glycoprotein; Lysosome; Zymogen.
 FT PROPEP 1 44 ACTIVATION PEPTIDE.
 FT CHAIN 45 390 CATHEPSIN D.
 FT ACT_SITE 77 77
 FT ACT_SITE 273 273
 FT DISULFID 71 140
 FT DISULFID 90 97
 FT DISULFID 264 268
 FT DISULFID 307 344
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 390 AA; 42488 MW; 5B38AA1C33C48D35 CRC64;

Query Match 11.8%; Score 314.5; DB 1; Length 390;
 Best Local Similarity 28.0%; Pred. No. 8.4e-18;
 Matches 113; Conservative 72; Mismatches 128; Indels 91; Gaps 21;

CC Tyr-26 bonds in the B chain of insulin.
 CC -!- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
 CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
 CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
 CC HORMONES AND RELATED SUBSTANCES.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 DR PIR; B38302; B38302.
 DR HSSP; P00791; 1PSA.
 DR MEROPS; A01.001; -.
 DR InterPro; IPR001969; Aspprotease_site.
 DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPsin.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Zymogen; Signal;
 KW Phosphorylation; Multigene family.
 FT SIGNAL 1 15
 FT PROPEP 16 59 ACTIVATION PEPTIDE.
 FT CHAIN 60 387 PEPsin II-1.
 FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
 FT ACT_SITE 93 93 BY SIMILARITY.
 FT ACT_SITE 276 276 BY SIMILARITY.
 FT DISULFID 106 111 BY SIMILARITY.
 FT DISULFID 267 271 BY SIMILARITY.
 FT DISULFID 310 343 BY SIMILARITY.
 SQ SEQUENCE 387 AA; 42070 MW; A6EC48F715541A48 CRC64;

Query Match 11.6%; Score 309; DB 1; Length 387;
 Best Local Similarity 27.1%; Pred. No. 2.3e-17;
 Matches 98; Conservative 68; Mismatches 130; Indels 66; Gaps 15;

Qy 75 YYVEMTVGSPPQTLNILVDTGSSNFAVG----AAPHPLHRYRQRLSSTYRDLRKGVYV 130
 |: :::|:| | :: ||||| | :: |||: : |||: : : :
 Db 75 YFGTISIGTPPQEFTVIFDTGSSNLWVPSTYCSSLACFLHKRFNPDDSTFQATSETLSI 134
 Qy 131 PYTQGWEGELGTDLVSIHPGNVTVRANIAAITESD---KFFINGSNWEGILGLAYAEI 187
 | | | | | : | : : : : | : : : ||||| |
 Db 135 TYGTGSMTGILGYDTVKV---GNIEDTNQIFGLSKTEPGITFLV--APFDGILGLAYPSI 189
 Qy 188 ARPDDSLPFFDSLVLKQTHV-PNLSLHLGAGFPLNQSEVLASVGGSMIIGGIDHSLYT 246
 : | : | | : : | : |||: | | : : ||| | ||
 Db 190 SASDAT--PVFDNMWNEGLVSEDLFSVYLSSNG-----EKGSMVMFGGIDSSYYT 237
 Qy 247 GSLWYTPIRREWYEVIIIVRVEINGQDLKM--DCKEYNYDKSIVDSGTTNLRPLPKKVFEA 304
 ||| : | : | | : : : |||: : | : : |||: | |
 Db 238 GSLNWVPVSHEGYWQITMDSITINGETIACADSC-----QAVVDTGTSLLAGPTSASK 291
 Qy 305 AVKSIKAASSTEKFPDGFVLGEQLV-CWQAGTTPWNIFFVISLYLMGEVTNQSFRTILP 363
 | | : : ||| : : | : | : | ||
 Db 292 IQSYIGASKNL-----LGENIISCSAIDSLPDIVF-----TINN 325
 Qy 364 QQYLRPVED-VATSQDDC---YKFAISQSSTGT--VMGAVIMEGFYVVFDRARKRIGFAV 417
 || | : ||| : : | | : : ||| : : | |
 Db 326 VQYPLPASAYILKEDDDCLSGFDGMNLDTSYGELWILGDVFIRQYFTVFDRAANNQVGLAA 385
 Qy 418 SA 419
 :|

RESULT 8

PEP4_MACFU

ID PEP4_MACFU STANDARD; PRT; 388 AA.
AC P27678;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Pepsin A-4 precursor (EC 3.4.23.1) (Pepsin I/II).
GN PGA.
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9543;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 16-70.
RC TISSUE=Gastric mucosa;
RX MEDLINE=92037645; PubMed=1935977;
RA Kageyama T., Tanabe K., Koiwai O.;
RT "Development-dependent expression of isozymogens of monkey
RT pepsinogens and structural differences between them.";
RL Eur. J. Biochem. 202:205-215(1991).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: hydrophobic, preferably
CC aromatic, residues in P1 and P1' positions. Cleaves 1-Phe-|-Val-2,
CC 4-Gln-|-His-5, 13-Glu-|-Ala-14, 14-Ala-|-Leu-15, 15-Leu-|-Tyr-16,
CC 16-Tyr-|-Leu-17, 23-Gly-|-Phe-24, 24-Phe-|-Phe-25 and 25-Phe-|-
CC Tyr-26 bonds in the B chain of insulin.
CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -!- MISCELLANEOUS: EACH PEPSINOGEN IS CONVERTED TO CORRESPONDING
CC PEPSIN AT PH 2.0 IN PART AS A RESULT OF THE RELEASE OF A 47 AA
CC ACTIVATION SEGMENT AND IN PART AS A RESULT OF STEPWISE PROTEOLYTIC
CC CLEAVAGE VIA AN INTERMEDIATE FORM(S).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X59753; CAA42425.1; -.
DR PIR; S19682; S19682.
DR HSSP; P00790; 1PSN.
DR MEROPS; A01.001; -.
DR InterPro; IPR001969; Aspprotease_site.
DR InterPro; IPR001461; AspproteaseA1.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.

DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Zymogen; Multigene family;
 KW Signal; Glycoprotein.
 FT SIGNAL 1 15 BY SIMILARITY.
 FT PROPEP 16 38 ACTIVATION PEPTIDE.
 FT PROPEP 39 62 ACTIVATION PEPTIDE.
 FT CHAIN 63 388 PEPSIN A-4.
 FT ACT_SITE 94 94 BY SIMILARITY.
 FT ACT_SITE 277 277 BY SIMILARITY.
 FT DISULFID 107 112 BY SIMILARITY.
 FT DISULFID 268 272 BY SIMILARITY.
 FT DISULFID 311 344 BY SIMILARITY.
 FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .).
 SQ SEQUENCE 388 AA; 41955 MW; A2923AB1F7FCDEB9 CRC64;

Query Match 11.5%; Score 307.5; DB 1; Length 388;
 Best Local Similarity 27.6%; Pred. No. 3.1e-17;
 Matches 108; Conservative 65; Mismatches 135; Indels 83; Gaps 17;

Qy 44 PRETDEEPEEPGRGGSFVEMVDNLRGKSGQGYVEMTVGSPPQTLNILVDTGSSNFAVGA 103
 | ||:| | :::: | : : :| | : : ||||| |
 Db 60 PTLIDEQPLE-----NYLDV-----EYFGTIGIGTPAQNFTVVFDTGSSNLWV-- 102

Qy 104 APHPFL-----HRYYQRQLSSTYRDLRKG VYVPYTQGWEGELGTDLVSI PHGPNVTV 156
 | : | : | |||| | | : | | || | | : : :
 Db 103 -PSVYCYSLACMDHNLFPQDSSSTYRATSKTVSITYGTGSMTGILGYDTVKV---GGISD 158

Qy 157 RANIAAITESDK-FFINGSNWEGILGLAYAEIARPDDSLEPFDSL VKQTHV-PNLFSLH 214
 | :||: ||: : :||| || | : | ||: | | :||: :
 Db 159 TNQIFGLSETEPGFFLYFAPFDGILGLAYPSIS--SSGATPVFDNIWNQRLVSQDLFSVY 216

Qy 215 LCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSLWYTPIRREWYVEVIIVRVEINGQDL 274
 | :|| | :| |||| | |||| : | : | | : : : : : : : :
 Db 217 LSAD----DQS-----GSVVIFGGIDSSYYTGSLNWVPVSVVEGYWQISVDSITMNGKTI 266

Qy 275 --KMDCKEYNYDKSIVDSGTTNLRLPKKVFEEAAVKSIIKAASSTEKFPDGEFWLGEQLV-CW 331
 | :|||:| | | | | : : : | | :| |
 Db 267 ACAKGC-----QAIVDTGTSLLTGPTSPIANIQSDIGASENSD-----GEMVVSCS 312

Qy 332 QAGTTPWNIFPVISLYLMGEVTNQSFRTILPQQY-LRPVEDVATSQDDCYK-----FAI 385
 : | :| || || | | : || |
 Db 313 AISSLPDIVF-----TINGVQYPLPPSAYILQSQGSCTSGFQGM DVP 354

Qy 386 SQSSTGTVMGAVIMEGFYVVFDRARKRIGFA 416
 :| :| | : : |||| :| |
 Db 355 TESGELWILGDVFIRQYFTVFD RANNQVGLA 385

RESULT 9

PEPA_CHICK

ID PEPA_CHICK STANDARD; PRT; 367 AA.
 AC P00793;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Pepsin A precursor (EC 3.4.23.1).
 OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=84004412; PubMed=6617663;
 RA Baudys M., Kostka V.;
 RT "Covalent structure of chicken pepsinogen.";
 RL Eur. J. Biochem. 136:89-99(1983).
 CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
 CC ALSO CLEAVED TO SOME EXTENT.
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: hydrophobic, preferably
 CC aromatic, residues in P1 and P1' positions. Cleaves 1-Phe-|-Val-2,
 CC 4-Gln-|-His-5, 13-Glu-|-Ala-14, 14-Ala-|-Leu-15, 15-Leu-|-Tyr-16,
 CC 16-Tyr-|-Leu-17, 23-Gly-|-Phe-24, 24-Phe-|-Phe-25 and 25-Phe-|-
 CC Tyr-26 bonds in the B chain of insulin.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 DR HSSP; P00794; 4CMS.
 DR MEROPS; A01.UPW; -.
 DR InterPro; IPR001969; Aspprotease_site.
 DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPSIN.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Zymogen; Glycoprotein.
 FT PROPEP 1 42 ACTIVATION PEPTIDE.
 FT CHAIN 43 367 PEPSIN A.
 FT ACT_SITE 77 77
 FT ACT_SITE 260 260
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .).
 FT DISULFID 90 95
 FT DISULFID 251 255
 FT DISULFID 290 323
 SQ SEQUENCE 367 AA; 40431 MW; 0C547E7FD8F5B341 CRC64;

Query Match 11.4%; Score 305; DB 1; Length 367;
 Best Local Similarity 24.0%; Pred. No. 4.5e-17;
 Matches 88; Conservative 70; Mismatches 124; Indels 84; Gaps 13;

Qy 75 YYVEMTVGSPPQTLNILVDTGSSNFAVGAAPHPFL-----HRYYQRQLSSTYRDLRKG 127
 || ::||:| | ::| ||||| | | : | : |||| :
 Db 59 YYGTISIGTPQQDFSVIFDTGSSNLWV---PSIYCKSSACSNHKRFDPSKSSTYVSTNET 115
 Qy 128 VYVPTYQGKWEGLGTDLVSIHPGPNVTVRANIAAITESDK-FFINGSNWEGILGLAYAE 186
 ||:| | ||| |:: ::|:| ::|:: | |::| ||||:|
 Db 116 VYIAYGTGMSGILGYDTVAV---SSIDVQNQIFGLSETEPGSFFYYCNFDGILGLAFPS 172
 Qy 187 IARPDDSLEPFFDSL VKQTHV-PNLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLY 245
 |: | ||::| | :||::| | | ::||| :
 Db 173 IS--SSGATPVFDNMMSQHLVAQDLFSVYLSKDG-----ETGSFVLFGGIDPNYT 220
 Qy 246 TGSLWYTPIRREWYYEVIIVRVEINGQDLK--MDCKEYNYDKSIVDSGTTNLRPLPKKVFE 303
 | ::|:| | |::: || : : : | :|||:| | :|: :
 Db 221 TKGIYWVPLSAETYWQITMDRVTVG NKYVACFFTC-----QAIVDTGTSLLVMPQGAYN 274

```

Qy      304 AAVKSIKAASSTE-----KFPDGFVLGEQLVCWQAGTTPWNIFPVISLYLMGEVTNQS 356
      :| : :| |      | ||      :: :: |
Db      275 RIIKDLGVSSDGEISCDDISKLPD-----VTFHINGHA----- 307

Qy      357 FRITILPQQYLRPVEDVATSQDDCYKFAISQSSTGT-----VMGAVIMEGFYVVFDRAR 410
      : |      ::|      | |      ::| | : :||:||||
Db      308 -----FTLPASAYVLNEDGSCMLGFENMGTPTELGEQWILGDVFIREYYVIFDRAN 358

Qy      411 KRIGFA 416
      ::| :
Db      359 NKVGLS 364

```

RESULT 10

PEPE_CHICK

```

ID  PEPE_CHICK      STANDARD;      PRT;      383 AA.
AC  P16476;
DT  01-AUG-1990 (Rel. 15, Created)
DT  01-AUG-1990 (Rel. 15, Last sequence update)
DT  28-FEB-2003 (Rel. 41, Last annotation update)
DE  Embryonic pepsinogen precursor (EC 3.4.23.-).
OS  Gallus gallus (Chicken).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC  Gallus.
OX  NCBI_TaxID=9031;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=88227903; PubMed=3131317;
RA  Hayashi K., Agata K., Mochii M., Yasugi S., Eguchi G., Mizuno T.;
RT  "Molecular cloning and the nucleotide sequence of cDNA for embryonic
RT  chicken pepsinogen: phylogenetic relationship with prochymosin.";
RL  J. Biochem. 103:290-296(1988).
CC  -!- DEVELOPMENTAL STAGE: SPECIFICALLY SECRETED DURING THE EMBRYONIC
CC  PERIOD IN THE CHICKEN PROVENTRICULUS (GLANDULAR STOMACH).
CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; D00215; BAA00153.1; -.
DR  PIR; A41443; A41443.
DR  HSSP; P00794; 4CMS.
DR  MEROPS; A01.028; -.
DR  InterPro; IPR001969; Aspprotease_site.
DR  InterPro; IPR001461; AspproteaseA1.
DR  Pfam; PF00026; asp; 1.
DR  PRINTS; PR00792; PEPSIN.
DR  PROSITE; PS00141; ASP_PROTEASE; 2.
KW  Hydrolase; Aspartyl protease; Digestion; Signal; Glycoprotein.
FT  SIGNAL      1      16      POTENTIAL.
FT  CHAIN      17      383      EMBRYONIC PEPSINOGEN.

```


RA Azuma T., Pals G., Mohandas T.K., Couvreur J.M., Taggart R.T.;
 RT "Human gastric cathepsin E. Predicted sequence, localization to
 RT chromosome 1, and sequence homology with other aspartic
 RT proteinases.";
 RL J. Biol. Chem. 264:16748-16753(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92112877; PubMed=1370478;
 RA Azuma T., Liu W.G., Vander Laan D.J., Bowcock A.M., Taggart R.T.;
 RT "Human gastric cathepsin E gene. Multiple transcripts result from
 RT alternative polyadenylation of the primary transcripts of a single
 RT gene locus at 1q31-q32.";
 RL J. Biol. Chem. 267:1609-1614(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Tatnell P.J., Kay J.;
 RT "HUMAN procathepsin E.";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
 RN [4]
 RP SEQUENCE OF 54-68; 77-95; 141-155; 275-285 AND 389-396.
 RX MEDLINE=90241267; PubMed=2334440;
 RA Athauda S.B.P., Matsuzaki O., Kgeyama T., Takahashi K.;
 RT "Structural evidence for two isozymic forms and the carbohydrate
 RT attachment site of human gastric cathepsin E.";
 RL Biochem. Biophys. Res. Commun. 168:878-885(1990).
 CC -!- FUNCTION: DUE OT ITS INTRACELLULAR LOCATION AND DISTRIBUTION IN
 CC LYMPHOID ASSOCIATED TISSUE, IT MAY HAVE A ROLE IN IMMUNE FUNCTION.
 CC -!- CATALYTIC ACTIVITY: Similar to cathepsin D, but slightly broader
 CC specificity.
 CC -!- SUBUNIT: Homodimer; disulfide-linked.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M84424; AAA52300.1; -.
 DR EMBL; M84413; AAA52300.1; JOINED.
 DR EMBL; M84417; AAA52300.1; JOINED.
 DR EMBL; M84418; AAA52300.1; JOINED.
 DR EMBL; M84419; AAA52300.1; JOINED.
 DR EMBL; M84420; AAA52300.1; JOINED.
 DR EMBL; M84421; AAA52300.1; JOINED.
 DR EMBL; M84422; AAA52300.1; JOINED.
 DR EMBL; J05036; AAA52130.1; -.
 DR EMBL; AJ250717; CAB82850.1; -.
 DR PIR; A42038; A34401.
 DR PDB; 1LCG; 17-APR-02.
 DR MEROPS; A01.010; -.
 DR Genew; HGNC:2530; CTSE.
 DR MIM; 116890; -.
 DR GO; GO:0007586; P:digestion; TAS.
 DR InterPro; IPR001969; Aspprotease_site.

DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPSIN.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal;
 KW Polymorphism; Pyrrolidone carboxylic acid; 3D-structure.
 FT SIGNAL 1 17
 FT PROPEP 18 53 ACTIVATION PEPTIDE.
 FT CHAIN 54 396 CATHEPSIN E.
 FT MOD_RES 18 18 PYRROLIDONE CARBOXYLIC ACID.
 FT ACT_SITE 96 96 BY SIMILARITY.
 FT ACT_SITE 281 281 BY SIMILARITY.
 FT DISULFID 60 60 INTERCHAIN (PROBABLE).
 FT DISULFID 109 114 BY SIMILARITY.
 FT DISULFID 272 276 BY SIMILARITY.
 FT DISULFID 314 351 BY SIMILARITY.
 FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 220 220 O-LINKED (POTENTIAL).
 FT CARBOHYD 333 333 O-LINKED (POTENTIAL).
 FT VARIANT 324 324 T -> I (IN dbSNP:6503).
 FT /FTid=VAR_014572.
 SQ SEQUENCE 396 AA; 42793 MW; 40B643C5FB01521E CRC64;

Query Match 11.3%; Score 301.5; DB 1; Length 396;
 Best Local Similarity 25.8%; Pred. No. 9.6e-17;
 Matches 100; Conservative 68; Mismatches 144; Indels 75; Gaps 16;

Qy 48 DEEPEEPGRRGSFVEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFAVGA---- 103
 |: :|| : :| |: ::||| :| :||| :| :
 Db 63 DQSAKEP-----LINYLD-----MEYFGTISIGSPPQNFTVIFDTGSSNLWVPSVYCT 110
 Qy 104 APHPFLHRYYQRQLSSTYRDLRKGVYVPYTQGWEGELGTDLVSIHPGNVTVRANIAAI 163
 :| | :| ||| : :| | | :| ||| :| : :
 Db 111 SPACKTHSRFQPSQSSTYSQPGQSFSIQYGTGSLSGIIGADQVSV-EGLTVVGQQFGESV 169
 Qy 164 TESDKFFINGSNWEGILGLAYAEIARPDDSLPFFDSLQVTHVP-NLFSHLHLCGAGFPL 222
 || :|:: : :||| | :| :| :| :| :| :| :
 Db 170 TEPGQTFVD-AEFDGILGLGYPSLA--VGGVTPVFDNMMAQNLVDLPMFSVYM----- 219
 Qy 223 NQSEVLASVGGSMIIGGIDHSLYTGSWYTPIRREWYVEVIIVRVEINGQDLKMDCKEYN 282
 | :| || || :||| :| : :| : : :| :| :| :
 Db 220 -SSNPEGGAGSELIFGGYDHSFSGSLNWVPVTKQAYWQIALDNIQVGG--TVMFCSE-- 274
 Qy 283 YDKSIVDSGTTNLRPLPKVFEEAAVKSIIKAASSTEKFPDGFVLGEQLVCWQAGTTPWNIFP 342
 :|||:| : :| : :| || || :| :| :| :| :
 Db 275 GCQAIVDTGTSLITGPSDKIKQLQNAIGAAP-----VDGEYAVE-----CANLNVMP 321
 Qy 343 VISLYLMGEVTNQSFRTILPQQYLRPVEDVATSQDDCYKFAISQSSTG----- 391
 : :| :| :| :| :| :| :| :| :| :| :| :
 Db 322 DVTFTING-----VPYTLSPYAT--TLLDFVDGMQFC-----SSGFQGLDIHPPAG 365
 Qy 392 --TVMGAVIMEGFYVVFDRARKRIGFA 416
 :| :| :| || ||| :| :| :| :| :| :| :
 Db 366 PLWILGDVFIRQFYSVFDRGNRRVGLA 392

RESULT 12

CATD_HUMAN

ID CATD_HUMAN STANDARD; PRT; 412 AA.
AC P07339;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Cathepsin D precursor (EC 3.4.23.5).
GN CTSD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85270436; PubMed=3927292;
RA Faust P.L., Kornfeld S., Chirgwin J.M.;
RT "Cloning and sequence analysis of cDNA for human cathepsin D.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4910-4914(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=87231068; PubMed=3588310;
RA Westley B.R., May F.E.B.;
RT "Oestrogen regulates cathepsin D mRNA levels in oestrogen responsive
RT human breast cancer cells.";
RL Nucleic Acids Res. 15:3773-3786(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=91299158; PubMed=2069717;
RA Redecker B., Heckendorf B., Grosch H.W., Mersmann G., Hasilik A.;
RT "Molecular organization of the human cathepsin D gene.";
RL DNA Cell Biol. 10:423-431(1991).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP SEQUENCE OF 1-22 FROM N.A.

RX MEDLINE=94085791; PubMed=8262386;
 RA May F.E., Smith D.J., Westley B.R.;
 RT "The human cathepsin D-encoding gene is transcribed from an estrogen-
 RT regulated and a constitutive start point.";
 RL Gene 134:277-282(1993).
 RN [6]
 RP SEQUENCE OF 1-22 FROM N.A.
 RX MEDLINE=95021301; PubMed=7935485;
 RA Augereau P., Miralles F., Cavailles V., Gaudet C., Parker M.,
 RA Rochefort H.;
 RT "Characterization of the proximal estrogen-responsive element of
 RT human cathepsin D gene.";
 RL Mol. Endocrinol. 8:693-703(1994).
 RN [7]
 RP SEQUENCE OF 170-180.
 RC TISSUE=Liver;
 RA Hochstrasser D.F., Frutiger S., Paquet N., Bairoch A., Ravier F.,
 RA Pasquali C., Sanchez J.-C., Tissot J.-D., Bjellqvist B., Vargas R.,
 RA Appel R.D., Hughes G.J.;
 RL Submitted (JUN-1992) to the SWISS-PROT data bank.
 RN [8]
 RP VARIANT VAL-58.
 RX MEDLINE=20179010; PubMed=10716266;
 RA Papassotiropoulos A., Bagli M., Kurz A., Kornhuber J., Forstl H.,
 RA Maier W., Pauls J., Lautenschlager N., Heun R.;
 RT "A genetic variation of cathepsin D is a major risk factor for
 RT Alzheimer's disease.";
 RL Ann. Neurol. 47:399-403(2000).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (3 ANGSTROMS).
 RC TISSUE=Spleen;
 RX MEDLINE=93223670; PubMed=8467789;
 RA Metcalf P., Fusek M.;
 RT "Two crystal structures for cathepsin D: the lysosomal targeting
 RT signal and active site.";
 RL EMBO J. 12:1293-1302(1993).
 RN [10]
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RC TISSUE=Liver;
 RX MEDLINE=93342076; PubMed=8393577;
 RA Baldwin E.T., Bhat T.N., Gulnik S., Hosur M.V., Sowder R.C. II,
 RA Cachau R.E., Collins J., Silva A.M., Erickson J.W.;
 RT "Crystal structures of native and inhibited forms of human cathepsin
 RT D: implications for lysosomal targeting and drug design.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:6796-6800(1993).
 CC -!- FUNCTION: Acid protease active in intracellular protein breakdown.
 CC Involved in the pathogenesis of several diseases such as breast
 CC cancer and possibly Alzheimer's disease.
 CC -!- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,
 CC that of pepsin A. Does not cleave the 4-Gln-|-His-5 bond in B
 CC chain of insulin.
 CC -!- SUBUNIT: CONSISTS OF A LIGHT CHAIN AND A HEAVY CHAIN.
 CC -!- SUBCELLULAR LOCATION: Lysosomal.
 CC -!- POLYMORPHISM: The Val-58 allele is significantly overrepresented
 CC in demented patients (11.8%) compared with nondemented controls
 CC (4.9%). Carriers of the Val-58 allele have a 3.1-fold increased
 CC risk for developing AD than noncarriers.

```

CC      -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; M11233; AAB59529.1; -.
DR      EMBL; X05344; CAA28955.1; -.
DR      EMBL; M63138; AAA51922.1; -.
DR      EMBL; M63134; AAA51922.1; JOINED.
DR      EMBL; M63135; AAA51922.1; JOINED.
DR      EMBL; M63136; AAA51922.1; JOINED.
DR      EMBL; M63137; AAA51922.1; JOINED.
DR      EMBL; BC016320; AAH16320.1; -.
DR      EMBL; L12980; AAA16314.1; -.
DR      EMBL; S74689; AAD14156.1; -.
DR      EMBL; S52557; AAD13868.1; -.
DR      PIR; A25771; KHHUD.
DR      PDB; 1LYA; 31-JAN-94.
DR      PDB; 1LYB; 31-JAN-94.
DR      PDB; 1LYW; 22-JUL-99.
DR      MEROPS; A01.009; -.
DR      SWISS-2DPAGE; P07339; HUMAN.
DR      Siena-2DPAGE; P07339; -.
DR      Genew; HGNC:2529; CTSD.
DR      MIM; 116840; -.
DR      GO; GO:0004192; F:cathepsin D activity; TAS.
DR      InterPro; IPR001969; Aspprotease_site.
DR      InterPro; IPR001461; AspproteaseA1.
DR      Pfam; PF00026; asp; 1.
DR      PRINTS; PR00792; PEPSIN.
DR      PROSITE; PS00141; ASP_PROTEASE; 2.
KW      Hydrolase; Aspartyl protease; Glycoprotein; Lysosome; Signal; Zymogen;
KW      Polymorphism; Alzheimer's disease; 3D-structure.
FT      SIGNAL          1          18
FT      PROPEP          19          64      ACTIVATION PEPTIDE.
FT      CHAIN           65         412      CATHEPSIN D.
FT      CHAIN           65         161      CATHEPSIN D LIGHT CHAIN (PROBABLE).
FT      CHAIN          169         412      CATHEPSIN D HEAVY CHAIN (PROBABLE).
FT      ACT_SITE        97          97
FT      ACT_SITE       295         295
FT      DISULFID        91         160
FT      DISULFID       110         117
FT      DISULFID       286         290
FT      DISULFID       329         366
FT      CARBOHYD        134         134      N-LINKED (GLCNAC. . .).
FT      CARBOHYD       263         263      N-LINKED (GLCNAC. . .).
FT      VARIANT         58          58      A -> V (ASSOCIATED WITH INCREASED RISK IN
FT                                         AD; POSSIBLY INFLUENCES SECRETION AND
FT                                         INTRACELLULAR MATURATION; dbSNP:17571).
FT                                         /FTid=VAR_011621.
FT      STRAND          67          74
FT      TURN            75          77

```

FT	STRAND	78	85
FT	TURN	86	89
FT	STRAND	90	97
FT	TURN	98	99
FT	STRAND	103	107
FT	TURN	108	109
FT	TURN	112	113
FT	HELIX	115	118
FT	TURN	119	119
FT	STRAND	123	123
FT	HELIX	125	127
FT	TURN	129	130
FT	STRAND	132	141
FT	STRAND	146	158
FT	STRAND	172	184
FT	HELIX	188	192
FT	STRAND	197	200
FT	HELIX	204	206
FT	HELIX	208	210
FT	HELIX	214	220
FT	TURN	221	222
FT	STRAND	228	233
FT	STRAND	243	247
FT	TURN	248	248
FT	HELIX	252	254
FT	STRAND	255	263
FT	STRAND	267	267
FT	TURN	268	269
FT	STRAND	270	279
FT	TURN	280	281

Query Match 11.3%; Score 300.5; DB 1; Length 412;
 Best Local Similarity 26.9%; Pred. No. 1.2e-16;
 Matches 123; Conservative 68; Mismatches 170; Indels 97; Gaps 21;

Qy	5	LPWLLLWMGAGVLPAGHTQHGI	RLPLR-----SGLGGAPLGL-----RLP	44
		: : : :		
Db	7	LPLALCLLAA---PASAL---VRIPLHKFTSIRRTMSEVGGSVEDLIAKGPVSKYSQAVP		60
Qy	45	RETDEEPEEPGRRGSFVEMVDNLRGKSGQGYVEMTVGSPPTLNILVDTGSSNFAVGAA		104
		: : : : : : : : : : :		
Db	61	AVTE-----GPIPEVLKNYMDAQ---YYGEIGIGTPPQCFTVVFDTGSSNLWVPSI		108
Qy	105	PHPFL-----HRYYQRQLSSTYRDLRKGVPYPTQGWEGELGTDIVSIP-----		149
		: :		
Db	109	HCKLLDIACWIHHKYNSDKSSTYVKNGTSTFDIHYGSGSLSGYLSQDTSVPCQSASSASA		168
Qy	150	HGPNVTVRANIAAITESDKFFINGSNWEGILGLAYAEIARPDDSLEPFFDSLQTHV-P		208
		: : : : : : : :		
Db	169	LGGVKVERQVFGEATKQPGITFIAAKFDGILGMAYPRIS--VNNVLPVFDNLMQQKLVDQ		226
Qy	209	NLFSLHLCGAGFPLNQSEVLASVGGSMIIGGIDHSLYTGSWYTPIRREWYVEIIVRVE		268
		: : : : : : : : : :		
Db	227	NIFS FYL-----SRDPDAQPGGELMLGGTDSKYKGSLSYLNVTWKAYWQVHLDQVE		278
Qy	269	I-NGQDLKMDCKEYNYDKSIVDSGTTNRLRPPKVFEEAAVKSIIKAASSTEFDPDGFWLGEQ		327
		: : : : : : : : :		

Db 279 VASGLTL---CKE--GCEAIVDTGTSLMVGVPVDEVRELQKAIGAVPLIQ-----GEY 325

Qy 328 LV-CWQAGTTPWNIFFPVISLYLMGEVTNQSFRTILPQQYLRPVEDVATSQDDCYKFAIS 386
 :: | : | | : | | : : : : | : | | : | :
 Db 326 MIPCEKVST-----LPAILTKLGG----KGYKLS--PEDYTLKVSQAGKTL--CLSGFMG 372

Qy 387 Q-----SSTGTVMGAVIMEGFYVVFDRARKRIGFAVSA 419
 | : | | : : | | | | | : | | | : |
 Db 373 MDIPPPSGPLWILGDFVFIGRYYTVFDRDNNRVGFAEAA 410

RESULT 13

PEP2_RABIT

ID PEP2_RABIT STANDARD; PRT; 387 AA.
 AC P27821;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Pepsin II-2/3 precursor (EC 3.4.23.1) (Pepsin A).
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91009127; PubMed=2129536;
 RA Kageyama T., Tanabe K., Koiwai O.;
 RT "Structure and development of rabbit pepsinogens. Stage-specific
 RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
 RT gene expression during development.";
 RL J. Biol. Chem. 265:17031-17038(1990).
 CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
 CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
 CC ALSO CLEAVED TO SOME EXTENT.
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: hydrophobic, preferably
 CC aromatic, residues in P1 and P1' positions. Cleaves 1-Phe-|-Val-2,
 CC 4-Gln-|-His-5, 13-Glu-|-Ala-14, 14-Ala-|-Leu-15, 15-Leu-|-Tyr-16,
 CC 16-Tyr-|-Leu-17, 23-Gly-|-Phe-24, 24-Phe-|-Phe-25 and 25-Phe-|-
 CC Tyr-26 bonds in the B chain of insulin.
 CC -!- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
 CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
 CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
 CC HORMONES AND RELATED SUBSTANCES.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M59235; AAA85369.1; -.
 DR HSSP; P00790; 1PSN.
 DR MEROPS; A01.001; -.
 DR InterPro; IPR001969; Aspprotease_site.

DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPSIN.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Digestion; Zymogen; Signal;
 KW Phosphorylation; Multigene family.
 FT SIGNAL 1 15
 FT PROPEP 16 59 ACTIVATION PEPTIDE.
 FT CHAIN 60 387 PEPSIN II-2/3.
 FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
 FT ACT_SITE 93 93 BY SIMILARITY.
 FT ACT_SITE 276 276 BY SIMILARITY.
 FT DISULFID 106 111 BY SIMILARITY.
 FT DISULFID 267 271 BY SIMILARITY.
 FT DISULFID 310 343 BY SIMILARITY.
 SQ SEQUENCE 387 AA; 42100 MW; 66FC331A3DC75891 CRC64;

Query Match 11.2%; Score 299; DB 1; Length 387;
 Best Local Similarity 26.9%; Pred. No. 1.5e-16;
 Matches 97; Conservative 64; Mismatches 134; Indels 66; Gaps 13;

Qy 75 YYVEMTVGSPQTLNILVDTGSSNEFAVGAAPHPF-----LHRYYQRLSSTYRDLRKG 127
 |: ::|:| | :| | | | | | : | : : : | | | : :
 Db 75 YFGTISIGTPPQDFTVIFDTGSSNLWV---PSTYCSSLACALHKRFNPEDSSTYQGTSET 131
 Qy 128 VYVPYTQGWEGELGTDLVSIHPGNVTVRANIAAITESDKFFINGSNWEGILGLAYAEI 187
 : : | | | | | | : : : : | | : : | | | | | | |
 Db 132 LSITYGTGSMTGILGYDTVKVGSIEDTNQIFGLSKTEPSLTFLF--APFDGILGLAYPSI 189
 Qy 188 ARPDDSLEPFFDSL VKQTHV-PNLFSLHLCCGAGFPLNQSEVLASVGGSMIIGGIDHSLYT 246
 : | : | | : : : | : | | : | | : : | | | | | | |
 Db 190 SSSDAT--PVFDNMWNEGLVSQDLFSVYLSSDD-----EKGSLVMFGGIDSSYYT 237
 Qy 247 GSLWYTPIRREWYVEVIIVRVEINGQDLKM--DCKEYNYDKSIVDSGTTNLRPKKVFEA 304
 | | | : | : | | : : : | | | : : | : : | | : | :
 Db 238 GSLNWWVPVSYEGYWQITMDSVSINGETIACADSC-----QAIVDTGTSLLTGP----TS 287
 Qy 305 AVKSIKAASSTEKFPDGFWLGEQLV-CWQAGTTPWNIFPVISLYLMGEVTNQSFRTITLP 363
 | : : | : | | | : : | : | : | : | : | : | : |
 Db 288 AISNIQSYIGASK----NLLGENVISCSAIDSLPDIVF-----TING 325
 Qy 364 QQYLRPVEDVATSQDDCYKFAISQSSTGT-----VMGAVIMEGFYVVFDRARKRIGFAV 417
 | | | : | : | : : : | : : | : : | | | : : | |
 Db 326 IQYPLPASAYILKEDDDCTSGLEGMNVDTYTGELWILGDVFIQYFTVFDRAANNQLGLAA 385
 Qy 418 S 418
 :
 Db 386 A 386

RESULT 14

PEP4_RABIT

ID PEP4_RABIT STANDARD; PRT; 387 AA.
 AC P28713;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Pepsin II-4 precursor (EC 3.4.23.1) (Pepsin A).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91009127; PubMed=2129536;
RA Kageyama T., Tanabe K., Koiwai O.;
RT "Structure and development of rabbit pepsinogens. Stage-specific
RT zymogens, nucleotide sequences of cDNAs, molecular evolution, and
RT gene expression during development.";
RL J. Biol. Chem. 265:17031-17038(1990).
CC -!- FUNCTION: SHOWS PARTICULARLY BROAD SPECIFICITY; ALTHOUGH BONDS
CC INVOLVING PHENYLALANINE AND LEUCINE ARE PREFERRED, MANY OTHERS ARE
CC ALSO CLEAVED TO SOME EXTENT.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: hydrophobic, preferably
CC aromatic, residues in P1 and P1' positions. Cleaves 1-Phe-|-Val-2,
CC 4-Gln-|-His-5, 13-Glu-|-Ala-14, 14-Ala-|-Leu-15, 15-Leu-|-Tyr-16,
CC 16-Tyr-|-Leu-17, 23-Gly-|-Phe-24, 24-Phe-|-Phe-25 and 25-Phe-|-
CC Tyr-26 bonds in the B chain of insulin.
CC -!- DEVELOPMENTAL STAGE: PEPSINOGENS IN GROUP I, II, AND III WHERE
CC THE PREDOMINANT ZYMOGENS AT LATE POSTNATAL STAGE.
CC -!- MISCELLANEOUS: THE EXPRESSION OF PEPSINOGEN GENES IS REGULATED BY
CC HORMONES AND RELATED SUBSTANCES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
DR PIR; D38302; D38302.
DR HSSP; P00790; 1PSN.
DR MEROPS; A01.001; -.
DR InterPro; IPR001969; Aspprotease_site.
DR InterPro; IPR001461; AspproteaseA1.
DR Pfam; PF00026; asp; 1.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Digestion; Zymogen; Signal;
KW Phosphorylation; Multigene family.
FT SIGNAL 1 15
FT PROPEP 16 59 ACTIVATION PEPTIDE.
FT CHAIN 60 387 PEPSIN II-4.
FT MOD_RES 129 129 PHOSPHORYLATION (POTENTIAL).
FT ACT_SITE 93 93 BY SIMILARITY.
FT ACT_SITE 276 276 BY SIMILARITY.
FT DISULFID 106 111 BY SIMILARITY.
FT DISULFID 267 271 BY SIMILARITY.
FT DISULFID 310 343 BY SIMILARITY.
SQ SEQUENCE 387 AA; 42052 MW; 21ADD07782A89585 CRC64;

Query Match 11.2%; Score 298; DB 1; Length 387;
Best Local Similarity 26.1%; Pred. No. 1.8e-16;
Matches 97; Conservative 66; Mismatches 122; Indels 86; Gaps 14;

Qy 75 YYVEMTVGSPPQTLNILVDTGSSNFAVGAAPHPF-----LHRYYQRQLSSTYRDLRKG 127
|: :::|:| | :: | | | | | | : | : : | | | : :
Db 75 YFGTISIGTPPQDFTVIFDTGSSNLWV---PSTYCSSLACALHKRFPEDSSSTYQGTSET 131
Qy 128 VYVPTYQGWEGELGTDLVSIHPGPNVTVRANIAAITESDKFF-----INGSNWE 177
: : | | | | | | : : | : : : | : : :

Db 132 LSITYGTGSMTGILGYDTV-----KVGSIEDTNQIFGLSKTEPGLTFLFAPFD 179

Qy 178 GILGLAYAEIARPDDSLEPFFDSLVKQTHV-PNLFSLHLCGAGFPPLNQSEVLASVGGSMI 236
 ||||| | : | : | ||: : | : |||: | : :

Db 180 GILGLAYPSISSDAT--PVFDNMWNEGLVSQDLFSVYLSSDD-----EKGSLVM 227

Qy 237 IGGIDHSLYTGSLWYTPIRREWYYEVIIVRVEINGQDLKM--DCKEYNYDKSIVDSGTTN 294
 |||| | |||| : | : | |:: : | |||: : | : |||: ||:

Db 228 FGGIDSSYYTGSLNWVPVSYEGYWQITMDSVSINGETIACADSC-----QAIVDTGTSL 281

Qy 295 LRLPKKVFEAAVKSIIKAASSTEKFPDGFWLGEQLV-CWQAGTTPWNIFPVISLYLMGEVT 353
 | | : | : | : | | | : | : | : | :

Db 282 LTGP---TSAISNIQSYIGASK---NLLGENVISCSAIDSLPDIVF----- 321

Qy 354 NQSFRTILPQQYLRPVEDVATSQDDCYKFAISQSSTGT-----VMGAVIMEGFYVVFED 407
 || || | : | : : | : : | : | : : |||

Db 322 -----TINGIQYPLPASAYILKEDDDCTSGLEGMNVDTYTGELWILGDVFIRQYFTVFD 375

Qy 408 RARKRIGFAVS 418
 || : | | :

Db 376 RANNQLGLAAA 386

RESULT 15

CATD_RAT

ID CATD_RAT STANDARD; PRT; 407 AA.

AC P24268;

DT 01-MAR-1992 (Rel. 21, Created)

DT 01-MAR-1992 (Rel. 21, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cathepsin D precursor (EC 3.4.23.5).

GN CTSD.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=Sprague-Dawley; TISSUE=Pituitary;

RX MEDLINE=91057150; PubMed=2243802;

RA Birch N.P., Loh Y.P.;

RT "Cloning, sequence and expression of rat cathepsin D.;"

RL Nucleic Acids Res. 18:6445-6445(1990).

RN [2]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 65-74; 118-127 AND 165-174.

RC TISSUE=Liver;

RX MEDLINE=91354249; PubMed=1883350;

RA Fujita H., Tanaka Y., Noguchi Y., Kono A., Himeno M., Kato K.;

RT "Isolation and sequencing of a cDNA clone encoding rat liver

RT lysosomal cathepsin D and the structure of three forms of mature

RT enzymes.;"

RL Biochem. Biophys. Res. Commun. 179:190-196(1991).

RN [3]

RP SEQUENCE OF 134-170.

RX MEDLINE=89034127; PubMed=3182800;

RA Yonezawa S., Takahashi T., Wang X., Wong R.N.S., Hartsuck J.A.,

RA Tang J.;

RT "Structures at the proteolytic processing region of cathepsin D.";
 RL J. Biol. Chem. 263:16504-16511(1988).
 CC -!- FUNCTION: Acid protease active in intracellular protein breakdown.
 CC -!- CATALYTIC ACTIVITY: Specificity similar to, but narrower than,
 CC that of pepsin A. Does not cleave the 4-Gln-|-His-5 bond in B
 CC chain of insulin.
 CC -!- SUBUNIT: OCCURS AS A MIXTURE OF BOTH A SINGLE CHAIN FORM AND TWO
 CC TYPES OF TWO CHAIN (LIGHT AND HEAVY) FORMS.
 CC -!- SUBCELLULAR LOCATION: Lysosomal.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X54467; CAA38349.1; -.
 DR PIR; S13111; KHRTD.
 DR HSSP; P07339; 1LYB.
 DR MEROPS; A01.009; -.
 DR InterPro; IPR001969; Aspprotease_site.
 DR InterPro; IPR001461; AspproteaseA1.
 DR Pfam; PF00026; asp; 1.
 DR PRINTS; PR00792; PEPSIN.
 DR PROSITE; PS00141; ASP_PROTEASE; 2.
 KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Signal;
 KW Lysosome.
 FT SIGNAL 1 20 POTENTIAL.
 FT PROPEP 21 64 ACTIVATION PEPTIDE (POTENTIAL).
 FT CHAIN 65 407 CATHEPSIN D.
 FT CHAIN 65 164 CATHEPSIN D 12 kDa LIGHT CHAIN.
 FT CHAIN 165 407 CATHEPSIN D 30 kDa HEAVY CHAIN.
 FT CHAIN 65 117 CATHEPSIN D 9 kDa LIGHT CHAIN.
 FT CHAIN 118 407 CATHEPSIN D 34 kDa HEAVY CHAIN.
 FT ACT_SITE 97 97 BY SIMILARITY.
 FT ACT_SITE 290 290 BY SIMILARITY.
 FT DISULFID 91 160 BY SIMILARITY.
 FT DISULFID 110 117 BY SIMILARITY.
 FT DISULFID 281 285 BY SIMILARITY.
 FT DISULFID 324 361 BY SIMILARITY.
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 258 258 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 15 15 D -> A (IN REF. 2).
 FT CONFLICT 163 163 D -> T (IN REF. 3).
 FT CONFLICT 205 205 K -> N (IN REF. 2).
 FT CONFLICT 262 262 K -> N (IN REF. 2).
 SQ SEQUENCE 407 AA; 44680 MW; C423AD4104D95F84 CRC64;

Query Match 11.1%; Score 297; DB 1; Length 407;
 Best Local Similarity 26.1%; Pred. No. 2.3e-16;
 Matches 118; Conservative 76; Mismatches 170; Indels 88; Gaps 20;

Qy 6 PWLLLMGAGVLPAGTQHGIPLR-----SGLGGA--PLGLRLPRETDEEPEEP 54
 | :|| : |:| | : ||:| | : :||: | |:| |

Db 4 PGVLLLI-LGLLDASSAL-IRIPLRKFTSIRRTMTEVGGSVEDLILKGPI TKYSMQSSP 61
 Qy 55 GRRGSFVEMVDNLRGKSGQGYVEMTVGSPPQTLN ILVDTGSSNFAVGAAPHPFL----- 109
 : |:: | || |: |:| | : : | | | | | : |
 Db 62 RTKEPVSELLKNYLDAQ---YYGEIGIGTPPQCFTVVFD TGSSNLWVPSIHCKLLDIACW 118
 Qy 110 -HRYYQRQLSSTYRDLRKGVYVPYTQGWEGELGTDLVSI PHGPNVTVRANIAAITESDK 168
 | | | | | : | | | | | | | : : : : | :
 Db 119 VHHKYNSDKSSTYVKNGTSFDIHYGSGSLSGYLSQDTVSV P-----CKSDLGGIKVEKQ 172
 Qy 169 FF-----INGSNWEGILGLAYAEIARPD DSLEPFFDSL VKQTHV-PNLFSLHLCG 217
 | : : | | | : | | : : | | | : | | : | : | :
 Db 173 IFGEATKQPGVVFIAAKFDGILGMGYPFIS--VNKVL PVFDNLMKQKLVEKNIFS FYL-- 228
 Qy 218 AGFPLNQSEVLASVGGSMIIGGIDHSLYTGS LWYTPIRREWYVEVIIVRVEINGQDLKMD 277
 : || : : | | | | | : | : | : | : : | : :
 Db 229 -----NRDPTGQPGGELMLGGTDSRY YHGELSYLNVTRKAYWQVHMDQLEV-GSELT L- 280
 Qy 278 CKEYNYDKSIVDSGTTNLR LPKKVFEAAVKS IKAASSTEKFPDGFWLGEQLV-CWQAGTT 336
 || : : | | : | | : | | : | | : : | : :
 Db 281 CK--GGCEAIVDTGTSL LVGPVDEVKELQKAIGAVPLIQ-----GEYMI PCEKVSS- 329
 Qy 337 PWNIFPVISLYLMGEVTNQSF RITILPQQYLRPVEDVATSQDDCYKFAIS-----Q 387
 | : | : | | : : | : | : | : : | :
 Db 330 ----LPIITFKLGGQ-----NYELHPEKY I LKVSQAGKT-----ICLSGFMGMDIPPP 373
 Qy 388 SSTGTVMGAVIMEGFYVVFDRARKRIGFAVSA 419
 | : : | | : : | | | | | : | | | : |
 Db 374 SGPLWILGDVFIGCY YTVFDREYNRVGF AKAA 405

Search completed: January 21, 2004, 09:23:03
 Job time : 25.9063 secs